Efficient Multi-site Statistical Downscaling Model for Climate Change

A Thesis
Submitted in Partial Fulfillment of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

in

GIS and Remote Sensing

by

Maharana Pratap Singh
(Reg. no. 2012RGI01)

Under the Supervision of

Dr. Varun Singh

GIS Cell

MOTILAL NEHRU NATIONAL INSTITUTE OF TECHNOLOGY ALLAHABAD

PRAYAGRAJ–211004, INDIA

September 2018

Undertaking

The work presented in my thesis titled **"Efficient Multi-site Statistical Downscaling Model for Climate Change"**, submitted to the GIS Cell, Motilal Nehru National Institute of Technology Allahabad, for the partial fulfilment of the **Doctor of Philosophy** degree in **GIS and Remote Sensing**, is my original work. I have neither plagiarized nor submitted the same work for award of any other degree. In case this undertaking is found incorrect, I accept that my degree may be unconditionally withdrawn.

September 2018 (Maharana Pratap Singh)

Prayagraj 2012RGI01

Certificate

Certified that the work contained in the thesis entitled **“Efficient Multi-site Statistical Downscaling Model for Climate Change"** has been carried out under my supervision and this work has not been submitted elsewhere for a degree.

(Dr. Varun Singh)

Thesis supervisor

Acknowledgements

I am indebted to my supervisor Dr. Varun Singh. His expert guidance has been invaluable.

I acknowledge the financial support as research fellowship by the **Government of India, Ministry of Human Resource Development (MHRD)** Department of Higher Education. I also thank Motilal Nehru National Institute of Technology (MNNIT) Allahabad, India.

I wish to record my sincere thanks to the faculty members and staff members of GIS Cell, MNNIT Allahabad for their help and kind support during my research work. I also thank the Student Research Committee (SRC) members Prof. Shiv Datt Kumar and Prof. Suneeta Agarwal for their useful suggestions.

I am thankful to my Father and Sister for their endless support and love in various ways.

(Maharana Pratap Singh)

Dedication

This thesis is dedicated to the memory of my mother Anita Singh, a kind and caring mother whom I always miss.

Abstract

Climate change and impacts studies are gaining importance in wake of changing climate and its impact. Climate models namely General Circulation Models have been developed by different research groups to study the impact of climate at Global Scale and they are the primary dataset available for modelling global climate change in the future. However, owing to their coarse spatial resolution, GCM models are not appropriate for impact studies at local scale having the finer spatial resolution. Therefore, for impact studies, climate models available at global scale are correlated with atmospheric and climate conditions like temperature and precipitation at local scale through downscaling process. Different downscaling techniques ranging from simple to dynamic downscaling techniques have been developed by the researchers to develop the mathematical models that correlate the GCM outputs with local observations.

Among these downscaling techniques, statistical downscaling techniques are most widely used techniques owing to easy of its implementation through computer based tools. SDSM is one of the widely used software for statistical downscaling that utilizes statistical downscaling technique for downscaling the GCM data-set. However, the available statistical downscaling software tools are not appropriate to automate the downscaling process for multiple grids of a given area of interest (AOI). Using the existing downscaling tools, manual intervention is required to downscale the GCM data at local scale for large AOIs having the sizeable spatial extent.

In this research work, a novel generalized downscaling model namely Efficient Multi-site Statistical Downscaling Model (EMSDM) based on the multivariate regression technique has been developed to automate the downscaling process for multiple grids. EMSDM can be applied to automate the downscaling of GCM data to multiple local grids of a AOI. Internal procedures

of EMSDM are programmed in platform independent C programming language for efficiently handling large quantum of GCM and local observation data and carrying out the complex mathematical computations like inversion of large matrices. For demonstrating, the applicability of the model, GCM model namely second generation Canadian Earth System Model (CanESM2) (CanESM2) developed by the Canadian Centre for Climate Modelling and Analysis (CCCma) of Environment and Climate Change Canada and local daily precipitation and temperature data-set acquired from Indian Meteorological Department (IMD) have been used for carrying out downscaling using the proposed model. India has been selected as AOI.

On basis of analysis of downscaling results generated by the model, it can be concluded that proposed model can efficiently be used to carry out statistical downscaling the AOI (comprising of multiple grids) irrespective of its extent. Results generated by the proposed model can be utilized by investigators to carry out climate impacts studies for AOI having large spatial extent.

Moreover, in order to facilitate the spatial geo-visualization of downscaling results, a web GIS based framework has been developed to geo-visualize the time series data generated by EMSDM. In addition of the downscaling, EMSDM is able to generate valuable spatial data-set pertaining to local observation and GCM outputs of given area of interest. These spatial date-set can utilized by the decision makers to investigate spatial distribution of climatological parameters like temperature, precipitation etc.

Contents

List of Figures

List of Pseudo Codes

List of Tables

List of Acronyms & Symbols

The list of acronyms and symbols described below show several acronyms and symbols that will be appear within the thesis many.

Acronyms

CanESM2	The second generation Canadian Earth System Model
CCCma	Canadian Centre for Climate Modelling and Analysis
EMSDM	Efficient Multi-site Statistical Downscaling Model
GCM	General Circulation Model
GCM	Global Climate Model
RCM	Regional Climate Model
CMIP3	Coupled Model Intercomparison Project Phase 3
CMIP5	Coupled Model Intercomparison Project Phase 5
CRU	Climate Research Unit
ENSO	El Niño-Southern Oscillation
IMD	India Meteorological Department
IPCC	Intergovernmental Panel on Climate Change
LARS-WG	Long Ashton Research Station Weather Generator
NOAA	National Oceanic and Atmospheric Administration
SDSM	Statistical DownScaling Model
WKT	Well Known Text

Symbols

α_k	k^{th} Regression coefficient
σ	Standard deviations
N	Number of observations
r_s	Spearman's Rank Correlation Coefficient
$\boldsymbol{H}$	Hat Matrix
$\boldsymbol{X}$	Set of Predictors
$\boldsymbol{x}$	Predictor
$\boldsymbol{Y}$	Set of Predictands
$\boldsymbol{y}$	Predictand
$\mathbb{C}^*$	Set of non zero complex numbers
$\mathbb{X}$	Ranking of X
$\mathbb{Y}$	Ranking of Y
$\mathcal{C}$	Category
$^\circ$	Degree
$^\circ\mathbb{C}$	Degree Celsius (Unit of Temperature)
r	Pearson Correlation Coefficient

Chapter 1

Introduction

1.1 General

India is facing a dismal situation. Out of the twenty major river basins, fourteen are being considered over stressed due to the population explosion, over-exploitation of resources and other activities which in turn resulting in negative impact of climate. The negative impact of climate change has made the problem intricate. Water availability has decreased from 1816 cubic meters per capita to 1545 cubic meters in a span of ten years from 2001 to 2011. The temperature of the Indian subcontinent is expected to rise by 2.5°ℂ to 4.5°ℂ by 2100. Moreover, according to the Intergovernmental Panel on Climate Change (IPCC) Technical Report on Climate Change and water changes in the large-scale hydrological cycle have been related to an increase in the observed temperature and resulting in irregular precipitation pattern over number of decades. Henceforth, the overall net impact of climate change on water resources is negative.

In order to investigate the climate change of a region, climate models have been formulated and prediction has been carried out using models by applying the downscaling process and developing the trends for climate variables like temperature, precipitation etc. for future years or decades. Downscaling

is a technique that correlates large scale climatic variables like atmospheric pressure and geopotential height, to local surface variables, for example, temperature and precipitation. Although existing research work on climate and hydrology provides the methodology to carry out downscaling of climate data at finer scales, framework for automating the downscaling process of given area of interest (AOI) irrespective of its spatial extent is not available. While considering this fact, the objective of this thesis is to develop computational framework for downscaling of climate data for given AOI irrespective its spatial extent. In this chapter, a general introduction provides an overview on need, objectives, scope, research contribution and organization of the thesis.

1.2 Need of Research

Due to non-availability of proper computational framework for downscaling the global climate model at local level for large areas, aggregate climate prediction for large regions like country is an integrate task. More specifically, due to non-availability of automation framework, analyst requires to carry out the downscaling of climate data for large number of sub-regions of AOI manually. Hence, in this Ph.D. research work, a computational framework for automating the grid wise downscaling for AOI like a country or state of a country has been proposed. Proposed computational framework is implemented as a set of software modules that are combinedly is given name as Efficient Multi-site Statistical Downscaling Model (EMSDM). For subsequent discussion, developed framework is abbreviated as EMSDM.

1.3 Objectives

Following are the objectives of the research work:

1. Selection of suitable downscaling technique for downscaling.

2. Development of relevant data structure for GCM and local data-set as required for downscaling.
3. Generation of spatial database for grid wise downscaling for a selected study area.
4. Development of mathematical model for statistical downscaling.
5. Development and implementation of a statistical downscaling model.

1.4 Scope of Research

Present research work is focused towards the development of a computational framework for downscaling the climate data at regional scale viz. country or a state. Owing to ease of its implementation using computer programs, statistical downscaling technique using regression techniques have been adopted for downscaling. Statistical downscaling is a type of downscaling technique that correlates large scale climatic variables of Global Climate Model (GCM), to local climate variables. This relationship is developed using a statistical model which can then be used to generate future data for climate prediction [Wilby et al., 2004]. GCM viz. CanESM2 and precipitation data acquired from India Meteorological Department (IMD) are used for application of the EMSDM.

1.5 Research Contribution

Following are the two major contributions of the research work:

1. This contribution endeavours to provide design guidelines pertaining to organization and management of climate data viz. precipitation, temperature used for downscaling and subsequent analysis. In brevity, following representation reinforces the contribution meaning:

   $$\boxed{\textbf{Data} : \textbf{Organization} \Rightarrow \textbf{Management}}$$

2. Development of computational framework for analysis and information generation. This contribution provides the "value addition" to acquired

data for information dissemination to the analyst i.e. how data can be transformed into useful climate information for given AOI under the purview of proposed framework. In brevity, following representation reinforces the contribution meaning:

$$\boxed{\textbf{Information} \Rightarrow \textbf{Data} + \textbf{Value Added}}$$

1.6 Organization of the Thesis

In Chapter 1, a general overview, need, objectives, scope, research contribution and organization of the research are presented. In Chapter 2, the preliminaries (concepts) like Climate, Weather, Atmosphere etc. relevant to present research have been discussed in brief. These preliminaries form the basis for discussions in subsequent chapters. In Chapter 3, review of downscaling approaches and their types with major focus on statistical downscaling has been carried out. Review of existing research works pertaining to statistical downscaling has also been carried out. Chapter 4 discusses the mathematical background related to major concepts including correlation, multiple linear regression, multi-variate regression and model solution for developing the mathematical model based on regression techniques. In Chapter 5, "Efficient Multi-site Statistical Downscaling Model (EMSDM)" and its underlying steps are discussed in detail. Chapter 6 discusses the implementation of EMSDM in detail. This chapter discusses the pre-processing of the data-sets and necessary algorithms for implementing the steps discussed in Chapter 5. In Chapter 7, application of EMSDM for India using CanESM2 and IMD data-sets has been presented to demonstrate its applicability for the specified area of interest. Finally, in Chapter 8, the major conclusions drawn from the research and recommendations as an overview of further study and implementation related to the current research have been discussed.

Chapter 2

Preliminaries

2.1 General

Research on climate research draws its guidelines for various aspects. In this chapter, key underlying terminologies pertaining to climate research are discussed in brief. These preliminary terms form the basis for the discussion carried out in subsequent chapters.

2.2 Terminologies

2.2.1 Weather

Weather is the atmospheric phenomena caused by the transfer or movement of the energy. Transfer of energy in the atmosphere occurs via movement of air; which causes the major weather phenomena [Allaby, 2009; Pont, 2014]. Association between time, place and atmosphere signifies the Weather. In other words the condition of the Atmosphere at a particular time and place is known as Weather of that place. It can be defined in brevity as:

"Weather is a Spatio-temporal phenomenon of Atmosphere."

Weather is associated with the physical conditions in the atmosphere (humidity, temperature, air pressure, wind, and precipitation) that exists over short time scales, generally days or weeks. In general, major weather patterns

are the consequence of rotation of the Earth and non-uniform heating of the atmosphere due to its exposure to solar radiation [Saha, 2008]. Weathers patterns result in development of high and low pressures at a geo-location on the Earth. Major factors which are studied as the components of the weather of a particular location on Earth are [Bychkov et al., 2010; Yiğit, 2015]:

1. Temperature
2. Humidity
3. Precipitation
4. Cloudiness
5. Visibility
6. Wind
7. Atmospheric Pressure
8. Solar radiation

2.2.2 Climate

Climate signifies the general weather condition that exists for longer period of time. Therefore, climate is the average weather condition that is quantified for the given set of decade(s) or centuries. Climate can be defined as

> *" Climate is statistics of Spatio-temporal phenomenon of Atmosphere."*

While weather signifies the short term atmospheric condition in time and space, climate signifies the long term atmospheric condition. Mark Twin curtly differentiates the climate and the weather as:

> *" Climate is what one expect and weather is what one gets."*

Moreover Climate describes the trends of the weather and helps in predicting the weather condition of the future. Climate of a particular geo-location is shaped by factors like glacier, mountains, ocean currents, planetary motions, winds etc. [Allaby, 2009].

Statistically, the climate of a place is defined as a weather statistics that prevails for the long-term. In short, climate is "statistics of weather". Quantitatively measurement of Climate can be carried out by determining the long term statistics of various weather parameters like temperature and rainfall; which are also referred as climate elements. Climate of a particular place also affected by extreme weather events. These extreme weather events are not treated as anomaly but used to determine better future climate change projections.

A Climate study can be carried on different geographical scales. Due to the diverse topography of the Earth's surface, places of tens of kilometers or just a few kilometers across can have its own climate influence; which can be termed as local climate of that place. For example, an urban area has its own climate due to prevalence of urban heat island effect; which is easily differentiable from the surrounding area in context of climate. In brief, it can be stated that the urban heating is a local climatic phenomenon. For large geographical extent like individual state or individual country, climate is termed as regional climate. Regional climate represents weather statistics or pattern of particular region i.e. a state or a country. Moreover, global climate statistics can be derived from the regional climate statistics.

Different geographic location on the Earth is classified into different climate zones. Climate classification given by famous Russian climatologist Wladimir Koppen's climate classification remains the most commonly used system recognized by geographical and climatological societies across the world. Koppen has classified climate zones as desert, polar, temperature, tropical, and

subtropical climates [Allaby, 2009]. Climate zones exhibit variable changing patterns of precipitation and temperature. Holistically, climate researchers investigate the global climate.

2.2.2.1 Key determining factors of the climate

Table 2.1 enumerates the factors that signify the climate of the area or region

Table 2.1: Factors affecting the Climate

S.No.	Factors	Remark
1	Altitude	Higher altitude areas are colder than the low altitude areas.
2	Latitude	Areas near the equator are warmer.
3	Humidity	Areas near the water bodies like river or ocean have high humidity in comparison to desert areas.
4	Topography	Mountain often block the circulation of the air henceforth acts as a barrier for air effecting the weather.

As depicted in Table 2.1, there are different factors that affect the climate of a given area. The most important factor is latitude. Owing of the ellipsoidal shape of the Earth along with its flattening at poles, locations which are in close proximity of the equator are more exposed to the solar radiations in comparison to the locations that are nearer to the North and South Poles. This temperature dissimilarity between the locations controls the atmosphere and shifting the heat far apart from the equator in the direction of the poles. This general circulation of the atmosphere is divided into the different circulation

cells. These cell signifies disparate pressure zones and wind belts [Allaby, 2009].

Other components of climate comprises of land-sea distribution, mountains and oceans. Ocean currents have a profound influence on regional and global climate. The Gulf Stream in the Atlantic affects the weather of Northwest Europe. The periodic El Niño current in the equatorial Pacific can have negative effects on the weather of some areas of South America and Australasia. Coastal regions usually exposed to mild and humid maritime climates, while the internal regions of the Earth have more continental climates that experience exterme winters and hot summers. Mountains influence regional climate as well as local climates and moreover disrupts the circulations of winds [Rafferty, 2011; Shrestha et al., 2014].

2.2.3 Atmosphere

Layers of gases surrounding the Earth are collectively termed as Earth's atmosphere. The atmosphere just looks like one vast blanket of gases which surrounds the Earth. These layers of gases are held up together by Earth's gravitational forces. The atmosphere keeps the Earth warm with the help of solar radiations for sustainability of the life-form on the Earth. Further, it protects the life-form from harmful solar radiations. For water cycle on planet Earth's Atmosphere is a major component. Most importantly, it provides the breathing air for life-form to live. The major composition of the Earth's Atmosphere is given in Table 2.2[Alexander, 2012; Allaby, 2009]:

As observed from the Table 2.2, that Nitrogen and Oxygen makes up 99% of the atmosphere. The composition of the Greenhouse Gases (GHG) in the atmosphere is less than 0.05%; out of this, 0.04% of the total gaseous component is Carbon Dioxide (CO2).

Generally the atmosphere is classified into five layers, each layer has its own

Table 2.2: Earth's Atmosphere gaseous components percentage

Gas		**Volume in %**
Name	**Formula**	
Nitrogen	N_2	78
Oxygen	O_2	21
Argon	Ar	0.9
Other Gases Combined		0.1

characteristics and properties. Following are the atmospheric layers [Bychkov et al., 2010; Saha, 2008]:,

1. Troposphere
2. Stratosphere
3. Mesosphere
4. Thermosphere
5. Exosphere

Another classification of atmosphere in seven layers is given below [Bychkov et al., 2010]:

1. Troposphere
2. Stratosphere
3. Mesosphere
4. Chemosphere
5. Thermosphere
6. Ionosphere
7. Exosphere

Troposphere and Stratosphere are the lowest layers in both of the classifications. Troposphere and the lower part of Stratosphere where the ozone resides

are the part of atmosphere responsible for the weather; remaining parts of the atmosphere is does not affect the weather. Troposphere is the densest layer of Earth's Atmosphere that comprises of 80% of its mass. Occurrence of major weather related activities is attributed to Troposphere.

2.2.4 Climate Model

Climate Model is the mathematical representation to portray the interaction between matter and energy in various regions of atmosphere, land and ocean of the climate system. These mathematical representations are based on the fundamental laws of physics, fluid motion, and chemistry. When a climate model's coverage is global then they are known as Global Climate Models (GCMs). These climate models which are basically mathematical equations are solved for 3D grid of atmosphere with the help of high computing resources over a number of time steps. Figure 2.1 provides the insights of climate models. As given in Figure 2.1 each of 3-dimensional grid cells are the mathematical representation that describe the materials in it and the way energy moves through it. These mathematical representations are based on the fundamental laws of physics, fluid motion, and chemistry.

Climate Model's resolution is defined by grid cell size. Climate Model with higher level of details has smaller size of the grid cells. Climate Model with more details have more grid cells, so there is a need of more computing power to generate that Climate Model. Figure 2.2 depicts the typical spatial resolution used in climate models for four IPCC Assessment Reports. In the first Assessment Report (FAR) in 1990, major climate models used a grid cells of about 500 km. In the second Assessment Report (SAR) published in 1996, spatial resolution of climate model's grid cells was improved to 250 km. In the third Assessment Report (TAR) that is published in year 2001, grid cell size had reduced to about 180 km, while in Fourth Assessment Report (AR4)

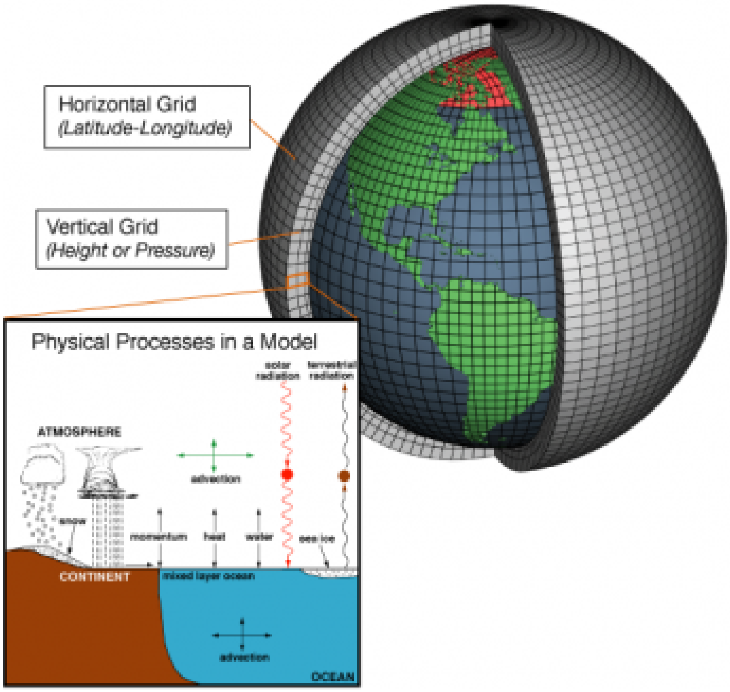


Figure 2.1: Conceptualization of Climate Model

climate models, grid size is reduced to 110 km. This improvement in spatial resolution allows such climate models to begin to make more accountable projections of regional climate in the future.

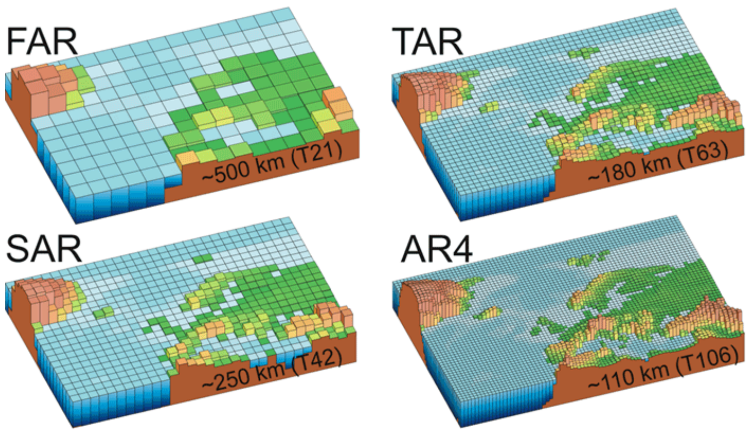


Figure 2.2: IPCC (AR4 WG 1 Chapter 1 page 113 Figure 1.4).

But as the coverage area is global and there are practical computing constraints, spatial resolution is about hundreds of kilometers. For further refining the coarse resolution of global climate model outputs to finer resolution climate

information Downscaling process is used, so that this finer climate information is used for regional and local topography with some better accountability.

A climate model is very similar to a weather forecast model, where weather forecast model is utilized for predicting short term atmospheric variability and change on the other hand climate model is for predicting long term atmospheric climate variability and change. Weather and realistic storms can be naturally simulated by climate models.

General Circulation Models, also called as Global Climate Models and abbreviated as GCMs. Global Climate Model are the mathematical formulation of the transfer of materials and energy through the climate system. Global Climate Models are based on well-documented physical processes. Some of input physical processes for the Global Climate Models are vegetation and soil wetness, glacial and sea ice, ocean circulations, human emissions of greenhouse gases and other pollutants, and some other inputs are wind direction, wind speed, air temperature, pressure, and humidity etc. Output of global climate model is known as climate projections. these are basically statistics and used to investigate the answer of climate related issues. Projections from the climate model are not expected to be exactly the same as those that really occur, but they should have the same typical characteristics.

2.2.5 Climate Change

Climate change is a change observed in the statistical distribution of weather patterns for a large period of time viz. decades to millions of years. Climate change signifies the change in average weather conditions, or in the time variation of weather within the perspective of longer-term average conditions. Climate change is attributed to factors like biotic processes, non-uniform exposure of the Earth to the solar radiation, plate tectonics and volcanic eruptions. Certain types of human activities are also attributed as major reasons

of recent climate change, that is also referred as global warming [Chen, 2012]. However their is no general concurrence available in existing literature that which specific term should be used to refer to anthropogenic forced changes in climate viz. whether Global Warming or Climate Change. Investigators in the area of climate studies are continuously improvising to analyse historical and future climate by using climate related observations and mathematical models. Comprehensive historical climate record based on different processes like geological evidence from borehole temperature profiles, analyses of sediment layers and stable-isotope etc., are continuously in development. GCMs based on the earth sciences are frequently utilized in theoretical approaches to synchronize the historical climate data-set, carry out future projections, and correlate causes and effects in climate change. Factors that can affect climate are termed as climate forcings or "forcing mechanisms. These are classified as internal or external mechanism. Internal forcing mechanisms are natural processes that are the components the climate system. For example, the thermohaline circulation. Exterior forcing factors can be either natural mechanics or anthropogenic. Anthropogenic mechanism are attributed to humans like increased emissions of greenhouse gases. Natural mechanisms are attributed to natural phenomena like changes in solar output, the earth's orbit, and volcano eruptions. Physical evidence to observe climate change comprises of different types parameters like temperature, precipitation etc. Global records of surface temperature are available since late 19th century. In past periods, most of the conspicuous indications are indirect-climatic changes and they are correlated with changes in proxy indicators that reflect climate like glacial geology, ice cores, vegetation and sea level change. Other physical evidence includes cloud cover over arctic sea, melting of ice of glaciers and precipitation etc. [Shrestha et al., 2014].

2.2.6 Climate Projection

A climate projection is the mathematically simulated trend of the climate system to a scenario of future concentration or emission of Green-House Gases (GHGs) and aerosols, which is derived using climate models. In common practice, a projection can be considered as any portrayal of the future and the pathway leading to it. However, a more explicit meaning that has been linked to the term "climate projection" by the IPCC which refers projection of model-driven estimations of future climate. Climate projections are differentiated from climate predictions so as to revel the fact that climate projections dependent on the factors like concentration, emission and radiative forcing scenario used etc. These scenarios are based on assumptions that may or may not be realized in future. Therefore, climate projections are affected by those uncertainties that are not related to the climate system.

2.2.7 Intergovernmental Panel on Climate Change (IPCC)

For assessment of climate change, the leading international body is Intergovernmental Panel on Climate Change (IPCC). IPCC was formed in 1988. IPCC was established by the World Meteorological Organization (WMO) and the United Nations Environment Programme (UNEP) and located at WMO headquarters in Geneva. IPCC as an intergovernmental body, opens its membership to all member countries of the United Nations (UN) and WMO. IPCC currently has 195 countries as members. In the IPCC's plenary Sessions governments of member countries participate for the review process. The IPCC Chair and Bureau Members are elected and work programme are framed during the plenary Sessions. The Secretariat coordinates all the IPCC work. The Secretariat also act as a link to assist communication between member Governments. The administration of IPCC is regularized in accordance to WMO and UN rules and procedures, including codes of conduct and ethical principles (as outlined in UN

Ethics, WMO Ethics Function, Staff Regulations and 2012/07-Retaliation).

As outlined in UN General Assembly Resolution 43/53 of 6 December 1988; the initial task for the IPCC was to prepare recommendations and comprehensive reviews with respect to the current scientific view and scientific facts of the science of climate change; potential impact of climate change on social, environmental and economical factors; formulating realistically possible response strategies and elements for adapting and mitigating risks of climate change in future.

IPCC does not monitor climate related data or parameter neither conduct any research on climate change. IPCC only assesses and reviews the latest scientific, technical, social, economical and environmental information compiled worldwide to understand the climate change. IPCC reports are drafted, reviewed, accepted, adopted and approved in these plenary Sessions. To ensure the completeness of the objectives and assessment of the current climate change information, IPCC adopt review as core part of its report generation and processing. To reflect wide range of expertise and views, IPCC accept the contribution of thousands of scientists from all over the world.

Decision makers and policy makers get benefited by balanced scientific information regardless of its geographical distribution form IPCC's reports due to its scientific and intergovernmental nature. By the recommendation of IPCC reports, governments and other authority form the policies based on their scientific content. IPCC reports are neutral with respect to policy, on other hand they can deal objectively with economical, environmental, scientific, social and technical factors relevant to the nature of particular policies.

IPCC First Assessment Report (AR1) of 1990 produced the scientific evidence emphasize the importance of climate change as a challenge requiring World Governments to tackle its effects. As an outcome of IPCC First Assessment Report, United Nations Framework Convention on Climate Change

(UNFCCC) is created. UNFCCC is the key international treaty to cope with the consequences of climate change and to reduce global warming. As a response of the UNFCCC, Special Reports, Methodology Reports and Technical Papers are compiled for this purpose. Kyoto Protocol in 1997 is the response of the IPCC Second Assessment Report (AR2) of 1995. The IPCC Third Assessment Report (AR3) was published in 2001 and the IPCC Fourth Assessment Report (AR4) was published in 2007. IPCC AR4 focuses significantly on the integration of climate change with sustainable development policies and relationships between mitigation and adaptation. Due to these efforts at the end of 2007, the IPCC was awarded the Nobel Peace Prize. IPCC Fifth Assessment Report (AR5) was published in 2014. IPCC Sixth Assessment Report (AR6) will be published in 2022, however the first order draft is completed and second order draft is in review.

The geographical distribution of participation of the scientific community has grown greatly for the work of the IPCC. Authors and contributors from all geographical locations involved in compiling and reviewing the reports. Geographical distribution of the topics covered by the reports has also grown significantly.

2.2.8 Downscaling

Downscaling is method adopted to deduce high-resolution information from low-resolution variables. This technique is based on dynamical or statistical approaches commonly used in different specializations like meteorology, climatology and remote sensing. Downscaling generally signifies refining spatial resolution, but it is sometimes also used for refining temporal resolution [Lloyd and Winsberg, 2018; Wilby and Wigley, 1997].

After obtaining the regression equation, the error and correlation between the target and the downscaled time series are obtained. Error and correlation

values provide overview to the analyst that how well regression equation represents the historical record. Finally the equation and similar set of predictors obtained from GCM are used to obtain historical condition and future projections downscaled to the desired AOI at the specified scale of spatial resolution (for example 0.5°). As downscaling is required to be carried out, it is imperative to carefully acquire a set of predictors from the GCM that adequately signifies the mathematical relation between the large scale atmosphere and the local scale climate parameters that are related to specified AOI. This selection is particularly become important where the topographical condition of AOI are very complex or there are some dominating predictor sets that affect AOI. For example, AOI located in mountainous or coastal areas. There can be large variations in model performance in these types of topographical regions [McGuffie and Henderson-Sellers, 2005].

Downscaling techniques are discussed in detail in Chapter 3 and Chapter 4. In chapter 3 downscaling and its classification have been discussed in detail. Chapter 4 provides the details of mathematical background related to the statistical downscaling method.

Chapter 3

Downscaling

3.1 What is Downscaling?

Downscaling is the process of linking two sets, one set with variables referring to large scale and another set with variables referring to small scale. In this research work, set of large-scale variable represent circulation pattern over a large spatial extent or region and set of variables referring the small scale can either of the climate variables like precipitation, temperature humidity etc. These measurements are the station measurement continuously collected from a given location during the specified time.

Large-scale variables have characteristic of varying smoothly and slowly in their spatial domain. Small-scale variable does not have smooth and slow varying characteristic. These small-scale variables are measured in the station in form of observations acquired through instruments like rain gauges, thermometers, barometers etc.

Using a downscaling approach, a real and physical mathematical relationship between these set of large-scale variables and set of small-scale variables are developed and evaluated. This mathematical relationship increases the reliability of climate projections and can be utilized for better projections and predictions pertaining to climate change.

Spatial scale selected for downscaling affects the refinement of climate variables. Spatial scale is large for the large-scale variables if they vary smoothly and slowly. Spatial scale is small if the climate-related changes that occurred over the small region are noticeable. Small-scale variable exhibits high covariance with a noteworthy component of the space, which are defined by the spatial coverage region of the large-scale variable, as the small-scale variable is contained within this space[Wilby et al., 2004; Wilby and Dawson, 2013; Wilby et al., 2002, 1998].

In other words, small-scale is referred to as local scale rather than the process which solely involves small spatial scales. Downscaling is only possible when large-spatial scale ought to be coupled with a local process in some manner. In absence of this coupling between variables, downscaling will produce non-interpretable outputs as climate projections. Hence, the coupling between the variables is the mandatorily required for the downscaling. As the coupling or the relation between the large-scale variable and the small-scale variables exhibit definitive relationship, downscaling produces more reliable climate predictions and better projection results pertaining to the climate change [Bhuvandas et al., 2014; Fowler et al., 2007].

To identify matching and harmonized time behavior on large-scale and small-scales, statistical downscaling approaches focus on the time dimension. Temporal variation mathematically represented as the function of time with a given time structure. The corollary of similar time structure on diverse spatial scales demonstrates the high temporal correlation. Downscaling can be carried out both in spatial and temporal contexts. Spatial and temporal contextual properties are moreover related to set of variables in which downscaling is performed [Bhuvandas et al., 2014; Ekström et al., 2015].

3.2 Why Downscaling?

Downscaling is used to make inference about some small-scale or local-scale climate variable at a given spatial location using a large-scale variable from global climate models. Without the downscaling, it is not relevant for real life or practical use to consider the global mean values from the large-scale variables derived from global climate models. For example, using downscaling approaches, investigators intend to apply global mean values from the large-scale variable to predict the precipitation at a local level or at smaller spatial extent. Though general circulation models (GCMs) characterize an essential role for studying climate change, however, it is not possible for general circulation models to give a practical useful prediction of the local climate. Consequently, it is desirable to downscale the climate projection from the general circulation models either through a high-resolution regional climate model (RCM) by the dynamic downscaling or through local station data by the statistical downscaling. Owing to following limitations of GCMs, downscaling is essentially be carried out to investigate the climate change at a local level:

1. Using the general circulation models, investigators are not able to carry out prediction of the real climate system at local scale, despite the fact that the general circulation models provide a reasonable prediction of the climate system on global scales.

2. Considering from the requirements of real-world problems, general circulation models are not suitable for providing high-resolution dataset for climate studies at a local scale. Currently, general circulation models datasets have low resolution and technical issues like artificial climate drift, atmosphere-ocean coupling, direct industrial effect and cloud representation etc. are not incorporated by them completely [Diaz-Nieto and Wilby, 2005; Haarsma et al., 2016].

3. Association of mathematical uncertainties with general circulation models and downscaling methods also exists. Moreover, these uncertainties interfere with the realistic prediction of climate change [Chen et al., 2011; Geerts B. and Linacre E., 1998; Hargreaves, 2010; Stouffer et al., 2017; Wilby et al., 1998].

4. Since the general circulation models have a coarse spatial resolution, they are incapable to signify aspects with spatial extent smaller than the spatial extent of general circulation models grid box size. Moreover, general circulation models are not capable to report for significant variations in the climate statistics inside a small region like the precipitation within the small cities or town. Through the application of downscaling, for smaller spatial extent, climate statistics with comparatively better reliability can be derived [Smerdon, 2012; Wilby et al., 2004; Wilby and Dawson, 2013].

3.3 Classification of Downscaling Methods

Downscaling is a broad study field in itself. There are numerous methods to classify downscaling methods. As the availability of large number of downscaling methods, there is a need of classification of these downscaling methods. Downscaling methods (approaches) can be classified into different ways. Generally downscaling methods can be classified on basis of applied mathematical concepts and usage. These methods as per the first classification are briefly discussed in the subsequent sections.

1. Simple Change Factors
2. Synthetic Statistical
3. Advanced (Deterministic Statistical)
4. Dyanamical

The second classification is based on basis of downscaling methods usage. Downscaling methods can be classified as per their usage. They are listed below:

1. **Simple Downscaling Methods**
   (a) Analogues
      i. Spatial Analogue
      ii. Temporal Analogue
   (b) Change Factors (Delta Method)
   (c) Bias Correction
      i. Local Scaling
      ii. Quantile-Quantile (QQ) Correction
   (d) Asynchronous Regional Regression Model (ARRM)
   (e) Bias Correction Constructed Analogues (BCCA)
   (f) Bias Correction Spatial Downscaling (BCSD)
   (g) Multivariate Adaptive Constructed Analogs (MACA)
2. **Advanced Downscaling Methods**
   (a) Statistical Downscaling
   (b) Dynamical Downscaling

3.4 Simple Downscaling Methods

Simple downscaling methods are the least complex and the least expensive methods of downscaling [Maraun and Widmann, 2018]. They are mainly been developed for generating higher resolution information from GCMs for developing impacts models. These techniques employ comparatively simplified transformation of the coarser outputs of GCMs, mainly temperature and precipitation [Maraun and Widmann, 2018]. Following subsections discuss the major types of simple downscaling methods.

3.4.1 Analogus Downscaling Methods

Analogues downscaling methods implement generation of approximate representation of the weather pattern for a specified day using the either one or number of histrorical weather patterns. Foremost implementation of ADM was carried out by Zorita and von Storch [1999] for downscaling the GCM data for Madrid, Spain. Analogues downscaling methods are applied in the case when climate models are not available. Investigator choose or select observation from some spatial location or some time period to reflect the climate change in the area of interest. Analogues methods can further be classified as:

3.4.1.1 Spatial Analogue

In spatial analogue methods, region having similar climatic condition to the area of interest is selected for downscaling. It is simple method. However, since it is difficult to find regions having similar climatic conditions, applicability of the method is limited.

3.4.1.2 Temporal Analogue

In temporal analogue methods, time period having desired climate is selected for downscaling. It is simple method. However methods are inflexible. However it is difficult to find time period with propers.

3.4.2 Change factors (Delta Method)

Change Delta method is one of the generally used downscaling method. It is most commonly used in investigation of the affects of climate change on water resources. In this method, changes in climate are estimated by carrying out comparative analysis of the future climate and current simulated climate derived through GCM and RCM models. Subsequently these changes are combined with observed local datasets having higher resolution. In general,

deviations in temperature (maximum and minimum) are augmented to the observed temperature values, and ratios of precipitation (future precipitation divided by the current precipitated) are combined with current observations using multiplication. In this approach, model biases are intrinsically corrected, since the scenarios are generated by altering observations. However, the corrections are only applicable to the mean of the climate change. Change factor method can be combined with the more complex stochastic methods.

3.4.3 Bias Correction

General circulation models have biases. Presence of the biasness in the model confines their direct utilization in impact studies. These biases can magnify since they may propagate by the impact models, finally resulting in simulation biases [Franzke et al., 2015; Haarsma et al., 2016; Stouffer et al., 2017]. Bias-correction methods can eliminate the systematic biases from the general circulation models, but they are unable to eradicate the unsystamatic biases applied to the general circulation models. Wilby et al. [2004] shows how mean standardization is corrected by bias-correction methods. Wood et al. [2004] shows how distribution in the quantile-based mapping is corrected by bias-correction methods. Likewise the general circulation models, regional climate models also have the biases [Maraun and Widmann, 2018; Shrestha et al., 2014].

3.4.3.1 Bias Correction (local scaling)

The bias correction with local scaling method intends to match the monthly mean of corrected values with that of observed values [Lenderink et al., 2007]. The method utilizes the monthly correction values calculated using the differences between observed and raw data for downscaling. Generally, precipitation values are corrected month wise using a multiplier and temperature with an

additive correction. Various point observations are taken to compute corrections. Subsequently, corrections have been applied on model outputs likewise rubber-sheeting process in order to obtain appropriate match for the output.

3.4.3.2 Bias Correction (Quantile-Quantile (Q-Q) Correction)

Quantile-Quantile correction method is similar to the local scaling method. However, in this method the entire distribution is corrected [Mearns et al., 2018]. The method was specially formulated to assist in the climate change studies. It intends to carry out correction on the distributions of the variables, hence possibly reducing the bias of extreme events estimators in climate model simulations [Cannon et al., 2015; Jeon et al., 2015; Quintana Seguí et al., 2010; Zhang et al., 2017]. The basic concept for quantile-quantile correction is discussed below [Scheuerer, 2018].

Let l be a location associated with a given analysis grid point and x be a location associated with given forecast grid point in the proximity of l. The basic objective of quantile mapping is to determine for each predicted f_x, to which quantile $q_{f,x}(k); k \in [0, 1]$ of the predicted local dataset it correlates with, and then map it to the corresponding quantile $q_{o,l}(k)$ of the observation dataset. The quantile functions $q_{f,x}$ and $q_{o,l}$ are approximated from the training sample; specifically investigator calculate the sample quantile $\hat{q}_{f,x}(n/100)$ and $\hat{q}_{o,s}(n/100)$ for $n \in \{1, 2, ..., 99\}$ and perform linear interpolation between these distinct values.

3.4.4 Bias Corrected Constructed Analogue (BCCA)

Bias Corrected Constructed Analogue (BCCA) method utilizes the historical climate observations to apply the correction. In BCCA, bias correction has been applied to GCM outputs. Subsequently, a set of similar 30 observed days will be selected on basis of comparison of bias corrected GCM future projection

with the historical observations. The similarity is reckoned on basis of spatial pattern and intensity of parameters like precipitation or temperature. A linear combination of the 30 selected days is developed to produce an analogue that nearly matches with GCM projections. The computed coefficients for the 30 historical days are applied to the local observations having the similar temporal resolution to generate a downscaled analogue. One major concern with this method is that owing to specific characteristics of GCM projections analogues may not exist for downscaling [Maurer et al., 2010; Walton et al., 2015].

3.4.5 Bias Correction Spatial Downscaling (BCSD)

Bias-corrected spatial disaggregation (BCSD) link the quartiles of the GCM outputs to historical local dataset patterns in order to generate daily time series and finally construct the downscaled grid. In more detail, BCSD comprises of trend removal, applying bias correction by linking link the quartiles of the GCM outputs to historical local dataset patterns and spatially resolving the local variables by interpolation of the bias-corrected anomalies and imposition of climatological means at finer scale [Hidalgo et al., 2008; Maurer et al., 2010; Mearns et al., 2018].

3.4.6 Asynchronous Regional Regression Model (ARRM)

Asynchronous Regional Regression Model (ARRM) is a comparatively complex statistical method that applies quantile regression to develop relationships between two quantities which are approximately normal distributed. These quartiles do not necessary have temporal correspondence, but should have similar statistical properties such as mean and variance. More specifically, ARRM uses piecewise regression to develop the relationship between observed and modelled quantiles and then downscale future projections [Mcginnis et al., 2014; Stoner et al., 2013]. ARRM at first, applies a sorting procedure to global

and local quantiles and map them against one another. Subsequently it finds, a six breakpoints between segments using linear regression over a moving window of fixed width to find points where the slope of the Q-Q map changes abruptly. Finally, it constructs a linear statistical model with the help of these breakpoints.

3.4.7 Multivariate Adaptive Constructed Analogs (MACA)

The Multivariate Adaptive Constructed Analogs (MACA) method is commonly used in wildfire applications [Abatzoglou and Brown, 2012]. Based on an observation based training dataset downscaling is performed by multivariate adaptive constructed analogs. Correction for biases in general circulation models output and spatially downscaling is performed by a multi-process approach. The latter is proficient uses the historical analogs from observational data that signifies various characteristics of regional climate [Abatzoglou and Brown, 2012].

3.5 Advanced Downscaling Methods

Researchers have classified the advanced downscaling as two major types viz. statistical downscaling and dynamical downscaling methods. Statistical downscaling (SD) consider the fact that regional climate in influenced by two key factors viz. (i) the large scale climatic state, and (ii) regional/local physio¬graphic features (e.g. topography, land-sea distribution and land use. From this point of view, regional or local climate information is extracted by first developing a statistical model that relates the large-scale climate variables (predictors) like surface air temperature and precipitation to regional and local variables (predictands). Subsequently, large-scale output of a GCM simulation is inputted into the developed statistical model to estimate the corresponding local and regional climate characteristics like rainfall, temperature etc. Statistical down-

scaling method and related research works are discussed in detail in section 3.10.

Dynamical downscaling approaches utilizes the regional climate models (RCMs), which generate finer resolution output on basis of relationships of atmospheric parameters for an AOI while taking GCM fields as boundary conditions. The physical consistency between RCMs and GCMs is controlled by the agreement of their large-scale circulations [Franzke et al., 2015]. Generalized steps of downscaling are shown in Figure 3.1. As conspicuous from Figure 3.1, dynamic and statistical downscaling processes are sometimes integrated together in order to obtain the optimal downscaling results in time and space [Hillel and Rosenzweig, 2011; Lloyd and Winsberg, 2018].

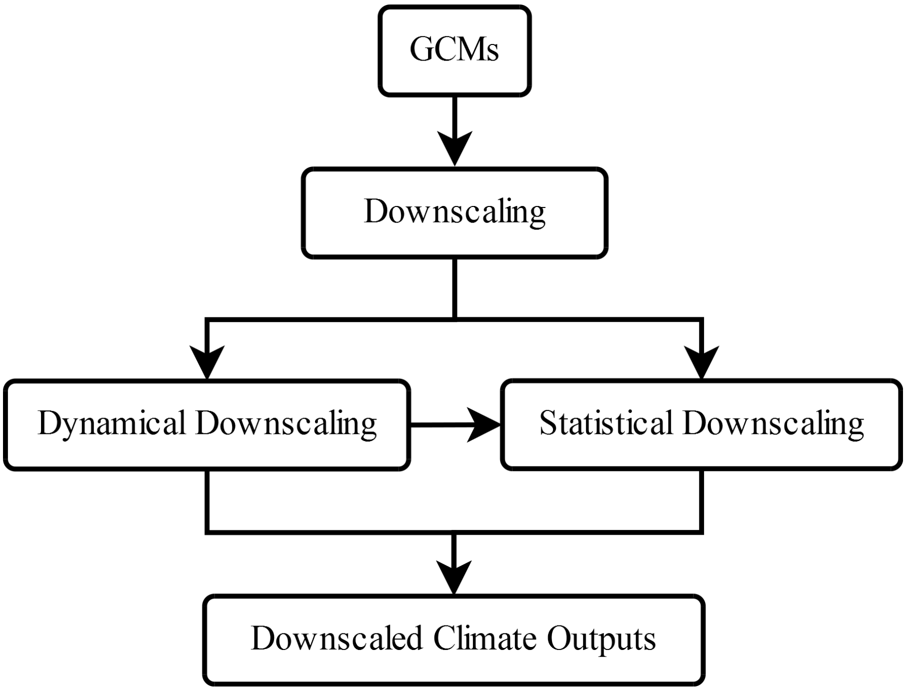


Figure 3.1: Generalized View for Downscaling Process

3.6 Qualities of Downscaling

Downscaling is a process having specific computation cost and time complexity associated with it. As this process has processing cost, so there must be some

qualities in the process and in the output produced by process. Some of these qualities of the downscaling process are given below :

Accuracy : It must be capable of reproducing high resolution historical records.

Feasibility : Downscaling methods must not be computationally complex.

Practical : Tractable to observations from immediate use in applied studies.

Products : It must be able to use multiple variables (predictors).

Resolution : Output climate projection with high spatial resolution ($<$10km).

Synchronicity : Downscaled weather fields adhere to the fundamental physical laws.

3.7 Data Used In Downscaling

As discussed earlier in this chapter downscaling is the process of linking two data-sets with a mathematical relationship viz. one referring to large scale variables and other referring to small scale variables. In this research work, set of large-scale variables represent the circulation pattern over a large spatial extent or region and set of small scale variables that represent the climate variables like precipitation, temperature humidity etc. for smaller spatial extent.

Large-scale variables are extracted from GCMs, as the out¬puts of these climate models. These large-scale variables represent the climate statistics for an area with a large geographical extent or large spatial extent. As these large-scale variables are the representation of the large spatial extent, these large-scale variables are come under the category of grid data.

Small-scale variables, which are the measurements from the climate station, represent the climate statistics for an area with small geographical extent or small spatial extent. As these small-scale variables are the representation of the small spatial extent, these small-scale variables are categorized as below: with some characteristics associated with it.

I. Gridded Data

- ✓ Spatially complete.
- ✓ Relies on gridded observations.
- ✓ Biases in obs → downscaled products

II. Point Observations (Station Data)

- ✓ Integrated into climate applications Currently.
- ✓ Eliminates biases in downscaled products.
- ✓ Subject to data availability and consistency.

3.8 Applied use of downscaled data

Sophistication of downscaling methods varies accordance to the need of end user application. The output of these downscaling methods is available as downscaled data. This downscaled data has the capability to reliably characterize the climate change of the given geographical extent. Downscaled data has many characteristics associated with them, some of the important characteristics are given below.

I. Data needs

- ✓ Spatial resolution.
- ✓ Temporal resolution.
- ✓ Multi-variable.

II. Vulnerability Assessment

- ✓ Extreme events.
- ✓ Inter-annual variability.
- ✓ Coincident sequencing.

III. Uncertainty Information

- ✓ Time horizon.
- ✓ Models used.

As downscaled data is the product of downscaling methods, selection of down-scaling methods is very important factor because not all downscaling methods are suitable for generating useful information as per the requirement of climate change studies.

3.9 Background Research Works

Most of the downscaling methods were developed on basis of conceptual and experimental developments pertaining to weather forecasting studies of the 1950s and 1960s. Klein et al. [1960, 1959, 1967] developed the perfect prognosis (PP) method to estimate the probability and type of precipitation, maximum, minimum and average temperatures, cloudiness and visibility at meteorological stations from numerical weather predictions. At first, authors generated the temporally synchronized statistical relationships between the climate parameters of interest and the observations of coarse-resolution climate parameters similar to weather forecasting models.

Subsequently, numerical model output was applied to the statistical relationships to estimate local weather at specified times. Finally, forecasting skill was subsequently used to gauge the efficacy of the developed model. For climatological applications, skill in forecasting measures the goodness of a forecast over a specified historical baseline of past observations. The forecasting approach may result in completely distinct skill measurements at different places, or maybe within the same place in various seasons. For example, spring weather can be driven by erratic local endemic conditions, whereas winter cold spur may correlate with discernible polar winds. The mean square

error, the coefficient of correlation between the forecasts and observations and nonsystematic bias in the forecast were some of the skill scores used in climate research [Murphy et al., 1989].

In continuation of above preliminary developments, Model Output Statistics (MOS), which is based on the statistical relationship between the local climate observations and output from the numerical model for specified projection time was developed and implemented by the researchers [Blocker, 1982; Glahn and Lowry, 1972; Klein and Hammons, 1975]. Thus, any biases in the numerical forecasts are compensated while applying the statistical scaling relationships. However, recalibration is considered necessary for each modification to existing, or development of new numerical forecast models. The foremost research work to transform GCM-scale output using PP approach is attributed to the seminal work carried out by Kim et al. [1984]. Authors used the monthly average surface temperature and monthly total precipitation at 49 meteorological stations in the State of Oregon to develop the mathematical relationship between GCM and local observations using empirical orthogonal functions (EOFs). The first EOF was able to elucidate the total variance of about 81% and 79% for the precipitation and temperature observations respectively. The importance of this seminal work was in the line of the fact that the prediction pertaining to local climate impacts can be feasible on basis of trends exhibited by the time-series of month wise weather anomalies observed at GCM grid-points that are superimposed over the area of interest.

Subsequently, the work of Kim et al. [1984] was advanced by Wigley, T. et al. [1990]. In addition to area average temperature and precipitation, authors used other predictor variables like geopotential heights, airflow gradients, the mean sea level pressure to downscale the GCM data to the local grid. On the basis of validation of their developed model using independent data, authors reported the spatial-mean explained variances range from 39% to 76%

for precipitation and from 58% to 87% for temperature. Authors asserted that most of the explained variance arises due averaging of predictand over the grid. They also demonstrated that site-specific changes can vary noticeably from those at the equivalent GCM grid-scale.

Karl et al. [1990] were foremost researchers who utilized GCM outputs to replicate the observed surface climate. They termed their approach as 'climatological projection by model statistics' (CPMS). They have used the primitive version of two-level atmospheric GCM developed by the Oregon State University for downscaling using PP and MOS approaches and local temperature, precipitation data acquired from five meteorological stations in USA. Twenty-two predictor variables acquired from this GCM were used to estimate daily cloud ceiling, precipitation and temperatures. They concluded that the MOS approach is preferable to PP approach since biases in the variances and mean of the GCM outputs can be removed by application of MOS. In continuation of the aforementioned research efforts, Von Storch et al. [1993] were the foremost researcher who had coined the term downscaling. They had employed PP approach for investigating Iberian rainfall anomalies in winter and sea level pressure field anomalies over the North Atlantic. Authors reported that the downscaled precipitation differed significantly from the GCM estimate of precipitation at the same locations.

Nearly at the same time, the preliminary RCMs were also developed by different researchers Giorgi [1990]; Giorgi and Bates [1989]. One of the earliest efforts towards the development of RCM using the nested regional modeling experiments. They were conducted in the western USA, where the intricate topography and coastline wield the significant influence over rainfall and temperature patterns. As reported by Giorgi et al. [1994] Pennsylvania State University/National Center for Atmospheric Research mesoscale model (MM4) having a spatial resolution of 60 km was tested using month-long winter time

simulations. MM4 was able to incorporate the topography of the region more realistically. In Europe, a similar model was used to carry out a double CO2 concentration experiment Giorgi et al. [1992]. Formative climate change investigations with RCMs were, however, hindered their efficacy to execute long simulations owing to limited computational facility, and non-availability of the historical GCM outputs at six-hourly frequencies. However, some of the mathematical principles governing the selection of domain size and resolution, as well as consideration of boundary forcing were soon developed [Giorgi and Mearns, 1991].

During the last two decades, numbers of research works have been reported in the area of downscaling of GCM outputs. Comprehensive review of downscaling methods have been also been carried out by different researchers [Fowler et al., 2007; Giorgi and Mearns, 1991; Hanssen-Bauer et al., 2005; Hewitson and Crane, 1996; Maraun et al., 2010; Von Storch et al., 1997; Wilby and Wigley, 1997; Xu, 1999].

Wilby and Wigley [1997] discussed the different downscaling approaches in detail. Authors investigated the latest downscaling approaches under four main classification namely limited-area climate models, regression methods, stochastic weather generators and weather pattern-based approaches. On basis of comparative analysis of these different approaches, authors reported that owing to feasibiliy of their implementation and less computing power requirenment, regression methods are preferred methods of downscaling More recently, Maraun et al. [2010] carried out a detailed review of the downscaling of precipitation data for climate change research. Barsugli et al. [2009] and Olsen and Gilroy [2012] asserted that downscaling techniques can generate fine resolution data at a local scale however, they will not correct large-scale errors in GCMs. Henceforth, advanced GCMs are in continuous development [García et al., 2014]. However, in spite of the substantial number of research works

in the area of downscaling, no agreement on the selection of an appropriate method for particular downscaling/bias-correction application has come out that result in some type of uncertainty in associated climate change impact investigations. As an example, Chen et al. [2011] reported the substantial uncertainty pertaining to the selection of downscaling methods. In the line of this problem, Puma [2012] developed the general guidelines for selecting a particular downscaling approach for climate scenario development on basis of complexity of analysis and spatiotemporal resolution (Figure 3.2).

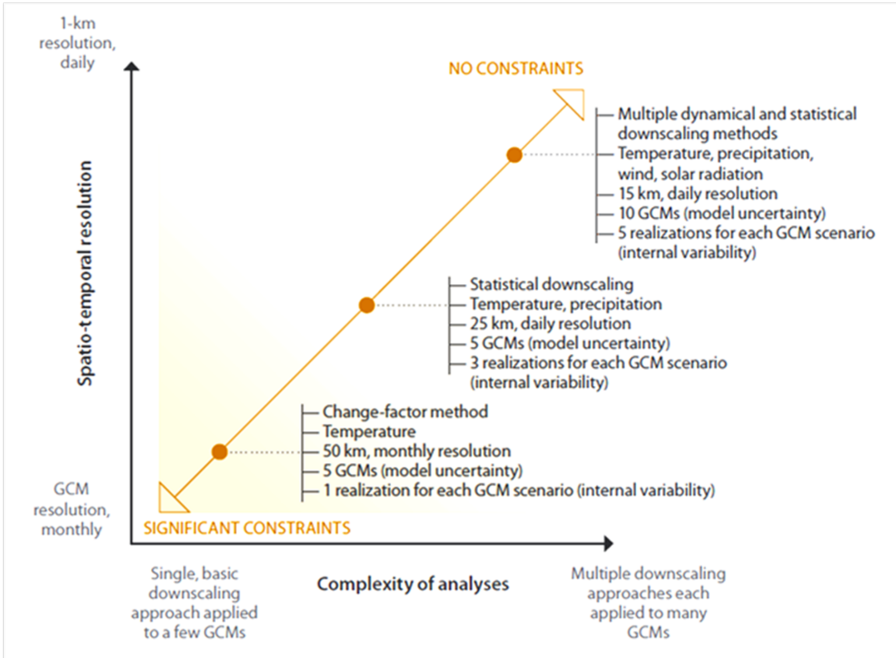


Figure 3.2: Selection of Downscaling Method Based on Spatio-Temporal Resolution [Puma, 2012]

3.10 Statistical Downscaling

3.10.1 Detailed Description

Statistical downscaling techniques usually ascribe to methods that statistically establish the relationship between larger scale atmospheric attributes from GCMs (predictors), such as greenhouse gas concentration, to local scale climate variable (generally point or grid estimates) (predictand), for example, monthly

temperature or precipitation. Moreover, owing to the facts that even Regional Climate Models (RCMs) do not have fine resolution to be used directly for impact studies and they are half biased, statistical downscaling is also required for RCMs as well. In addition to the aforementioned reasons, other major reasons of adopting statistical downscaling for downscaling RCM dataset are that it is computationally non-intensive and it facilitates generation of unbiased downscaled data.

There are various types of statistical downscaling that utilize different techniques, like artificial neural networks, weather generators, weather classification typing [Fowler et al., 2007; Giorgi et al., 1994]. Even though statistical downscaling methods vary both in terms of their intricacies and principles, they have two basic common processing steps viz. a training step and an application step. The training step uses two datasets. These two datasets are real-world local observations like precipitation or temperature and outputs of a physical model like RCM, GCM, reanalysis etc. Both sets are in synchronization with some past time period. Typically, the local observations will be of higher spatial resolution than the GCM outputs. During the training step, a mathematical relationship between these two datasets is developed through techniques like regression, forming a linkage between local observations and GCM outputs. These techniques represent linear or non-linear relationships between the local scale variable and large scale predictors. For example one has to obtain 0.5° precipitation data over the given area of Interest(AOI) using simple linear regression. At first, the sufficient historical observations from a weather station or a grid observation product from a normal area are acquired. Subsequently relevant predictors are acquired from historical reanalysis data like GCMs that are in sync with time with the historical observations of the station. Subsequently the regression equation is developed for. General expression of this equation is given in equation 3.1.

$$Y = c + \sum_{i=1}^{n} a_i X_i \tag{3.1}$$

where: Y = Historical Observations

c = Bias or Constant

n = Number of Parameters

a_i = ith Cofficient

x_i = ith GCM Parameter

After obtaining the regression equation, the error and correlation between the downscaled time series and the target are obtained. Error and correlation values provide the analyst to obtain the overview that how well regression equation represent the historical record. Finally the equation and similar set of predictors obtained from GCM are used to obtain historical condition and future projections downscaled to the desired AOI at the desired scale of spatial resolution (for example 0.5°). As downscaling is required to be carried out, it is imperative to carefully acquire a set of predictors from the GCM that adequately represent the relationship between the large scale atmosphere and the local scale climate variables related to specified AOI. This selection is particularly become important where the topographical conditions of AOI are very complex or there are some dominating predictor sets that affect AOI. For example, AOI located in mountainous or coastal areas. There can be large variations in model performance in these types of topographical regions [McGuffie and Henderson-Sellers, 2005].

In the application step, the derived mathematical relationship is applied to a set of GCM outputs pertaining to a different time period. The output of the application step is the generation of a dataset of surrogate observations, or downscaled results. In relation to climate change, the GCM outputs correspond to a specified future state for which different forcing parameters (e.g., greenhouse gases) have altered. The underlying concept behind the statistical downscaling for climate change investigation is to recalibrate the raw GCM

outputs for future climate state, imparting attributes of the observations during the historical (training) period, and in the process generating information at a higher spatial resolution as a substitute to the coarser information available at GCM level. Climate projections that have been advanced by statistical downscaling techniques are available as local observations at a fine resolution which generally are considered to be more suitable for use in many climate impact studies instead of the raw outputs of the GCM models.

Maraun et al. [2010] carried out further classification of the statistical downscaling methods into three types as prefect prognosis (PP), model output statistics (MOS) and weather generators. In PP, the statistical downscaling relationships are using local observations. In MOS, gridded RCM simulations and observations are utilized for developing the downscaling model. Weather generators are hybrid downscaling methods that use either PP/MOS approaches or both of them.

Pertaining to the types of statistical methods, downscaling can be classified as (i) categorical, (ii) continuous-valued and (iii) hybrid [Fowler et al., 2007; Wilby and Wigley, 1997]. In categorical downscaling, classifications and clustering approaches are used to develop statistical model between predictors and predictands [Zorita and von Storch, 1999]. In continuous-valued downscaling, models are developed using regression analysis in order to the map large scale predictors and local-scale predictands [Chandler and Wheater, 2002]. Mehrotra and Sharma [2010] carried out a non-parametric stepwise predictor identification analysis. In hybrid downscaling, various statistical approaches referred as weather generators are integrated together for downscaling Wilby et al. [2002].

Statistical downscaling model (SDSM) is widely used for statistical downscaling. It is developed by Wilby et al. [2002]. As per Wilby and Dawson [2013], SDSM is highly cited software in the climate research. Moreover, Long

Ashton Research Station Weather Generator (LARS-WG) is available to downscale weather data for a single site for the current and future climate [Semenov and Brooks, 1999]. LARS-WG is a proprietary software.

Large-scale climatic state and regional/local physiographic features (land use, land-sea distribution, topography etc.) are the two factors on which the regional climate depends. These fact about the regional climate is utilized by the statistical downscaling [von Storch and Navarra, 1995, 1999; von Storch and Zwiers, 1999; Zorita and von Storch, 1999]. From this point of view, small-scale or local climate statistics is derived by developing statistical downscaling models that relates the large-scale climate variables like precipitation and surface air temperature to small-scale or local climate variables. Subsequently, large-scale output of a GCM simulation is inputted into the developed statistical model to estimate the corresponding small-scale and local climate characteristics (climate statistics) pertaining like rainfall, surface air temperature etc.

Key advantages and disadvantages of statistical downscaling model are as described next.

I. Advantages

- ✓ Not computationally intensive.
- ✓ Applicable to GCM and RCM output.
- ✓ Provide Station/point values.

II. Disadvantages

- ✓ Lack of long/reliable observed series.
- ✓ affected by biases in the RCM and GCM.
- ✓ Not physically based e.g. climate feedbacks.
- ✓ Under-estimate variability and extremes.
- ✓ Assume stationary relationships in time.

3.10.2 Existing Research Works

During last few decades, different types of statistical downscaling approaches (methods) have been developed with a broad range of applications pertaining to climate change studies [Fowler et al., 2007; Giorgi et al., 2001; Hessami et al., 2008; Khalili et al., 2013; Lafon et al., 2013; Lee and Jeong, 2014; Lin et al., 2017; Orlowsky et al., 2010; Wilby et al., 2002]. Among different statistical downscaling methods, the regression approach is the most widely used statistical downscaling technique due to ease in its implementation and its relative less computational requirements in comparison to other methods [Tang et al., 2016a]. In the continuation aforementioned discussion, selected research works pertaining to statistical downscaling methods are discussed in the subsequent sections.

Khalili and Nguyen [2018] proposed a new statistical approach to the downscaling of daily maximum and temperature series for many different sites concurrently. Their approach is based on an integration of the modelling of the relation between local daily temperature extremes and the global climate predictors using regression method and modeling the stochastic component based using Singular-value decomposition (SVD) technique. Authors have used data of ten weather stations located in southeast region of Ontario and the southwest region of Quebec in Canada and two GCM datasets namely HADCM3 and CGCM3. They concluded that the proposed SVD-based SD procedure was tested and proven to be an effective tool for downscaling temperature extremes for many sites concurrently.

Serur and Sarma [2017] analysed the temperature and precipitation characteristics of Weyib River basin in Ethiopia in order to investigate the effect of climate change. They have used CanESM2 model for the RCP2.6, RCP4.5, and RCP8.5 scenarios. They have the developed the statistical downscaling model using the observed daily data of 12 meteorological stations.

Chen et al. [2018] developed the software namely Generator for Point Climate Change (GPCC) to generate the daily time series of climate change scenarios for local and AOI related climate change studies while utilizing the monthly climate projections from RCMs and GCMs for a particular grid. GPCC downscales monthly projections of GCMs or RCMs in a grid box to daily weather series at a point scale or station. and forecasting the minimum temperature of Raipur City. They had developed the statistically downscaled model for predicting the temperature for three future periods using Canadian Global Climate Model (CGCM) predictors for A1B and A2 climate forcing conditions. Authors have used Statistical Downscaling Model (SDSM) software using k-fold validation technique for forecasting the temperature for three different 15 years of time frame viz. FP-1 (2020-2035), FP-2 (2046-2064), and FP-3 (2081-2100). They reports CGCM model parameters viz. specific humidity at 850 hpa (nceps850gl), 500 hpa geopotential height (ncepp500gl), and surface airflow strength (ncep_fgl) were the most appropriate parameters to predicting future scenarios. Moreover they reported that comparison of mean monthly minimum temperature of generated scenarios with base period resulted in 1.1-11.2% increase in minimum temperature for A1B climate forcing condition.

Jaiswal et al. [2018] applied the statistical downscaling technique for downscaling and forecasting the minimum temperature of Raipur City. They had developed the statistically downscaled model for predicting the temperature for three future periods using Canadian Global Climate Model (CGCM) predictors for A1B and A2 climate forcing conditions. Authors have used Statistical Downscaling Model (SDSM) software using k-fold validation technique for forecasting the temperature for three different 15 years of time frame viz. FP-1 (2020-2035), FP-2 (2046-2064), and FP-3 (2081-2100). They reports CGCM model parameters viz. specific humidity at 850 hpa (nceps850gl), 500

hpa geopotential height (ncepp500gl), and surface airflow strength (ncep_fgl) were the most appropriate parameters to predict future scenarios. Moreover they reported that comparison of mean monthly minimum temperature of generated scenarios with base period resulted in 1.1-11.2% increase in minimum temperature for A1B climate forcing condition.

Yhang et al. [2017] implemented an integrated approach comprising of Dynamic and Statistical Downscaling approaches. They have tested their approach on East Asian summer monsoon precipitation data in order to obtain finely resolved data. The data-set acquired by the researchers covers a period of more than 57 years for monsoon Asia, the Middle East, and northern Eurasia. It is available on $0.5° \times 0.5°$ and $0.25° \times 0.25°$ grid mesh. They concluded that the integrated downscaling approach produced the better results in time and space in comparison to individual approaches. Smid and Costa [2017] carried out detailed review of downscaling approaches with application for impact studies in urban areas. They concluded that regression methods are the most suitable methods for climate change studies. AM et al. [2017] have developed a R-package namely spdownscale for statistical downscaling the climate data using quantile-quantile bias correction technique.

Pahlavan et al. [2017] proposed improved statistical model based on multiple linear regression (MLR) for carrying out statistical downscaling of monthly precipitation. They have termed their model as Monthly Statistical DownScaling Model (MSDSM). Moreover authors developed MSDSM based upon the general structure of Statistical DownScaling Model (SDSM). In order to enhance the efficacy of the model, they have adopted statistical modifications that incorporate bias correction using variance correction factor (VCF) to refine the pattern of computed variance. Authors demonstrated the efficacy of MSDSM through its application to 288 rain gauge stations scattered in different climatic zones of Iran. They concluded that MSDSM is comparatively

more suitable for reproducing the long-term mean and variance of monthly precipitation in comparison to SDSM.

Stennett-Brown et al. [2017] utilized the Statistical Downscaling Model (SDSM) to study the future projections of daily rainfall extremes for 39 stations and minimum and maximum temperature extremes for 45 stations in Caribbean and neighbouring regions. They have the reported that their developed models proved to be suitable for predicting the monthly mean daily temperatures and the frequencies of warm and cool days and nights between years 1961-2001. Authors have predicted variations in warm (cool) days and nights by years 2071-2099 under the scenarios A2 and B2 in comparison to years 1961-1990. However their developed models for rainfall demonstrated lower ability generally to model the monthly climatic trends of mean of daily rainfall and the spatial distribution of the average of yearly maximum number of consecutive dry days and mean yearly count of days with daily rainfall exceeding 10mm.

Hanel et al. [2017] developed a R package namely "musica" that provides functionality for validating statistical downscaling methods at multiple time scales and other statistical downscaling methods. The musica package is used to verify simulated runoff trends. Authors have demonstrated that conventional downscaling approaches results in significant biases in simulated runoff at all time scales for a given AOI.

Sa'adi et al. [2017] carried out investigation of the variation patterns of the rainfall owing to climate change of Sarawak Region in Borneo Island using statistical downscaling of GCM projections. Authors have used observed rainfall data to downscale the future rainfall from ensembles of 20 GCMs of Coupled Model Intercomparison Project phase 5 (CMIP5) for four Representative Concentration Pathways (RCP) scenarios, namely, RCP2.6, RCP4.5, RCP6.0 and RCP8.5. Model Output Statistics (MOS) based downscaling models were de-

veloped using two data mining approaches known as Random Forest (RF) and Support Vector Machine (SVM). Authors reported that SVM is able to downscale all GCMs with normalized mean square error (NMSE) of 48.2-75.2 and skill score (SS) of 0.94-0.98 during validation. Authors concluded that rainfall is varying owing to the affect of the monsoon season.

Onyutha et al. [2016] carried out comparison of outputs of three statistical downscaling methods viz. simplified, change factor and advanced (wetQP) quantile-perturbation-based methods for future prediction of rainfall on basis of daily rainfall for the period 1961-2000 at nine meteorological stations in the basin of Lake Victoria in Eastern Africa. Authors have used 14 GCMs from CMIP3 from 07 GCMs from CMIP5 for comparative analysis. Authors reported that outputs from the three SD methods are suitable for simulating the patterns of monthly rainfall totals. Moreover, the projected changes of seasonal or annual rainfall totals from the evaluated approches exhibit nearly same pattern with subtle differences. However authors reported the conspicuous differences in the results from the Delta, simQP and wetQP approaches in relation time series of quantiles of rainfall.

Vu et al. [2016] used the artificial neural network technique to statistically downscale GCMs at meteorological site locations in Bangkok. For implementing their proposed approach, authors used the large-scale predictor variables derived from GCM dataset namely CGCM3, ECHAM5 MIROC medium resolution (medres) and MPI-ESM-MR and local daily precipitation data for the period 1980-2000. The predictors are first selected over different grid boxes surrounding Bangkok region and subsequently the screening of the predictors was carried out them using PCA in order to select suitable correlated predictors for ANN application. They concluded are statistically downscaling techniques are computationally inexpensive and can easily be applied to analyse the output data from different GCM dataset.

George et al. [2016] implemented a statistical downscaling using local polynomial regression for developing the future climate projections of rainfall in a selected catchment in Kerala, India. Authors have assessed their developed model in comparison to compared with MLR amd ANN models. Authors concluded that the local polynomial regression model performs better in forecasting the rainfall of the area of study.

Sigdel and Ma [2016] applied the SDSM for downscaling the precipitation data in the three climatic regions of Nepal. Authors carried the calibration of downscaling model using large-scale atmospheric analysis data provided by National Centers for Environmental Prediction (NCEP). For the validation of the model the outputs of downscaled scenarios A2 and B2 of the HadCM3 model has been utilized. On basis of validation of their developed model, authors reported that the average R2 value period was 0.84 and SDSM is suitable for simulating precipitation. Moreover, their calibrated model demonstrates the better performance over humid region in comparison to subhumid and arid regions.

Poggio and Gimona [2015] presented an approach to downscale climate models based on a combination of Generalised-additive-models and geostatistics. They have evaluated the effectiveness of integrated approach using monthly means of temperature and rainfall to predict soil moisture condition. The climate model data was downscaled using integrated methods combining generalized additive models (GAMs) and kriging in order to reproduce the spatial pattern. The downscaled climate model data were subsequently corrected for bias using interpolated ground station data (1961-1999). Authors had implemented their developed approach using open-source software namely GRASS and R.

Sachindra et al. [2014a,b] presented the two statistical downscaling models for a precipitation station located in Victoria, Australia. Authors developed

the first model on the basis of National Centers for Environmental Prediction/-National Center for Atmospheric Research (NCEP/NCAR) reanalysis outputs and the second model on basis of outputs of Hadley Centre Coupled Model version 3 GCM (HadCM3). Authors have applied the multi-linear regression (MLR) technique for development of the model for a precipitation station situated in Victoria, Australia. Authors reported that during both calibration and validation, these models under-predicted the high precipitation values and over-predicted near-zero precipitation values. Moreover, bias correction technique has been applied by the authors to correct the raw outputs of model developed using HadCM3.

Jha et al. [2013] presented a downscaling approach based on multiple-point geostatistics (MPS) to downscale climate variables viz. skin surface temperature, soil moisture, and latent heat flux. They have implemented the geostatistical approach of direct sampling (DS) based upon the MPS to sample spatial patterns within training images to develop the downscaling model at different spatial scales. On basis of data derived from a RCM of the Murray-Darling basin in southeast Australia with 50 and 10 km spatial resolutions, authors carried out assessment of their developed approach. Authors analysed that downscaling results and concluded that results from their developed approach are in good agreement with the spatial distribution of WRF reference variables at a finer spatial scale for the surface temperature and heat flux irrespective of seasons. However, on basis of analysis of downscaled results for soil moisture, authors concluded that DS approach is not suitable for reproducing soil moisture for small-scale land surface features, specifically for water bodies like canals, lakes, and reservoirs.

Wilby and Dawson [2013] evaluated the application of widely used software namely Statistical DownScaling Model (SDSM). Authors have discussed the underlining conceptual and technical evolution of SDSM. Moreover, they

carried out independent assessments of model capabilities of SDSM. On basis of review of related studies using SDSM, they concluded that SDSM provides reliable estimates of seasonal precipitation totals, extreme temperatures and areal and inter-site precipitation behavior. However SDSM produces less satisfactory results for frequency estimation of extreme precipitation amounts during summers. Further they reported that SDSM is unable to downscale simultaneously to multiple sites, however the basic model may be extended for downscaling for multiple sites through process of resampling [Wilby et al., 2003].

Guo et al. [2012] proposed an Automated statistical downscaling (ASD) method for carrying out the prediction of the daily precipitation of 138 meteorological stations in the Yangtze River basin for year 2010-2099 using regression based statistical downscaling approach. They have used HadCM3 model outputs of A2 and B2 scenarios. They concluded that their developed method was suitable simulating the amount and the change pattern of precipitation.

Raje and Mujumdar [2011] carried out comparison of outputs of three downscaling approaches viz. conditional random field (CRF), k-nearest neighbor (KNN), and support vector machine (SVM) for downscaling point-scale daily precipitation in the Punjab region, India, at six locations only for the monsoon regime. They concluded that CRF and KNN performed subtly better than SVM.

Ashiq et al. [2010] utilized the different interpolation models in GIS environment to generate fine scale ($250 \times 250 m^2$) precipitation surfaces from PRECIS precipitation data. Authors concluded that the multivariate extension model of ordinary kriging that utilizes the data is the appropriate model for downscaling the precipitation data during mansoon season.

Goyal and Ojha [2010] evaluated different linear regression-based downscaling models like forward, backward, direct, and step-wise regression for down-

scaling mean monthly precipitation in the arid Pichola watershed, India. They concluded that direct regression-based downscaling model performs better in comparison to other regression techniques for that investigated region.

Pagé et al. [2009] developed a software package namely dsclim is based upon the statistical downscaling methodology proposed by Boe and Terray [2008]. "dsclim" was developed on the basis of concepts of climate regime and large-scale circulations (LSC). Huth et al. [2008] compared the linear and non-linear statistical downscaling methods using daily temperature data of eight stations in Europe for winters. They have compared linear regression of grid-point values (point-wise regression), regression based on principal components analysis and artificial neural networks. They have compared the downscaling methods on basis of fit to observations, shape of the statistical distribution in terms of its skewness and kurtosis, temporal autocorrelations with 1 day lag, and interstation correlations. They concluded that pointwise linear regression outperforms all all the methods.

Hessami et al. [2008] developed an automated statistical downscaling methodology (ASD) using a backward stepwise regression for predictor's selection. Outputs from the third version of the coupled global Hadley Centre Climate Model (HadCM3) and first generation Canadian Coupled Global Climate Model (CGCM1) were used by the research to validate ASD over the period of 1961-1990 and downscaled models results are compared with observed temperature and precipitation from 10 meteorological stations of Environment Canada located in eastern Canada. Authors reported that in comparison to downscaling precipitation, ASD is more effective in downscaling the temperature.

Fealy and Sweeney [2007] proposed the downscaling technique based upon the generalized linear modeling approach that overcomes some of the difficulties encountered for prediction of daily precipitation. They have tested their

developed method to predict the precipitation amounts for a selection sites in Ireland.

Tripathi et al. [2006] formulated various SVM-based downscaling models in order to downscale monthly precipitation at various meteorological subdivisions (MSDs) in India. They concluded that the SVM-based model is more effective in downscaling the monthly precipitation in comparison to conventional ANN-based downscaling model.

3.11 Concluding Remarks

Global Climate Model (GCMs) can replicate the global and or continental-level of climate scenarios plausibly well. However, they still deficient in producing fine resolution dataset required for climate studies at local level (from example a river basin). Various approaches have been developed by the researchers in order to bridge the gap between climate and climate impact models. These approaches range from simple downscaling approaches like delta method to complex dynamic downscaling approaches that utilize the Regional Climate Models (RCMs). RCMs provide physical parameters or mathematical models at the regional scale. During last few decades, there has been a remarkable development in generation of RCMs and their capabilities to replicate the present-day climate at regional level. However factors like the inherent systematic errors that exist in GCMs dataset act as boundary conditions to RCMs, the high computational time, and the requirement for further downscaling for impact studies hamper the application of RCMs for downscaling the GCM data effectively [Wilby et al., 1998].

Simple downscaling methods ranging from simple scaling to intricate distribution mapping have been developed and implemented for downscaling in the last decade [Chen et al., 2012, 2013, 2011; Chen, 2012; Iizumi et al., 2011; Lafon et al., 2013; Mpelasoka and Chiew, 2009; Piani et al., 2010; Ryu et al.,

2009; Salvi et al., 2011; Sharma et al., 2007, 2009; Smitha et al., 2018; Terink et al., 2010; Teutschbein and Seibert, 2012; Teutschbein et al., 2011].

However, it is imperative to assess and carry out comparision of performances for a specific impact study. Chen et al. [2013] compared the performance of six bias correction methods for hydrological modeling for ten North American river basins. They concluded that bias correction methods performance depends upon the location and that a careful investigation must be carried out specifically for studies over the new areas of interest (AOI). Walton et al. [2017] concluded that both BCSD and BCCA do not perform satisfactory for downscaling temperature in the Sierra Nevada. Stoner et al. [2013] reported that simple downscaling methods do not yield satisfactory results for predicting extreme events pertaining to temperature and precipitation. Further they reported that simple methods are suitable for downscaling monthly/annual means, daily output for tropical regions. Regarding downscaled temperature and precipitation, the statistical approaches are comparatively more proficient in generating the temporal and spatial autocorrelation properties.

As the advancement of the simple methods, statistical downscaling methods based on the statistical relationship between large-scale climate parameters and local scale or regional level parameters like temperature or precipitation can be used. Such statistical methods are based on the argument that statistical relationships between the predictors and predictand remain tenable for future times as well. However, statistical downscaling can be used only at the locations where precipitation or temperature observations records are available at point and (or) grid levels.

On basis of comparative analysis carried out by the investigators, it can be concurred that statistical and dynamical downscaling methods exhibit similar performances [Murphy, 2000; Sun et al., 2015; Tang et al., 2016a; Walton et al., 2015; Wood et al., 2004; Yhang et al., 2017].

Murphy [2000]; Walton et al. [2015]; Wilby and Wigley [2000]; Wood et al. [2004] concluded that performance of statistical and the dynamical approaches are similar. Likewise, Wood et al. [2004], who investigated the six different downscaling methods in context of their effectiveness to simulate temperature and precipitation trends, did not observe any additional improvement for dynamical downscaling approaches. Wilby and Wigley [1997] discussed the merits and demerits of the statistical and dynamical downscaling techniques. They concluded that the dynamic downscaling method, which is based on a physically consistent process are able produce fine resolution data at local scale, however its efficacy relies upon the bias in GCM boundary conditions and the effect of regional forcing phenomena. Also, its use is restricted by its computational cost. Khan et al. [2006a] evaluated three statistical downscaling methods, Statistical DownScaling Models (SDSM), LARG-WG, and Artificial Neural Network. They compared their potential in terms of reproducing observations and concluded that SDSM proved to be the most appropriate method for downscaling. On basis of comparative analysis of six downscaling methods by Chen et al. [2012]; Chen [2012], the authors reported the apparent differences in projections derived from different downscaling methods. Smid and Costa [2017] carried out detailed review of downscaling approaches with application for impact studies in urban areas. They concluded that statistical downscaling approaches based on regression are the most suitable in context of climate change studies in urban areas.

In the next chapter, the mathematical background for implementing statistical downscaling based on regression analysis is discussed in detail.

Chapter 4

Mathematical Background

4.1 General

Steps associated with downscaling are based on mathematical concepts of correlation and regression analysis. Underlying steps of Efficient Multi-site Statistical Downscaling Model (EMSDM) implement these concepts for downscaling. In this chapter, these mathematical concepts are discussed in detail.

4.2 Predictor

In statistics, for the independent variable, a more contemporary term used for the variable is termed as predictor. In this research work, for statistical modelling large-scale variables used as input data are known as predictor, which described the circulation pattern over a spatial region. Contextually it is referred with many terminology like "input variable", "independent variable" or as "large-scale variable".

Mathematically, if it is a single series then it is referred as uni-variate and multi-variate in case of several parallel series. Mathematical symbology for uni-variate and multi-variate predictors are represented using the following notations as given in the table 4.1.

Table 4.1: Type of predictors and notation

Predictor Type	Mathematical Symbol
Uni-variate Predictor	$\vec{x}$
Multi-variate Predictor	$\vec{X}$

4.3 Predictand

In statistics, for the dependent variable, a more contemporary term used for the variable is termed as predictand. In this research for statistical modeling, small-scale variable used as output data is known as predictand, which represents the climate variable measured at climate station. Contextually it is referred with many term like "output variable", "dependent variable", "responding variable", "response variable", "small-scale variable" or as "regressand". A very common definition used in many places for it is like "That which is to be predicted".

In mathematical equation it is the left hand side term of the equation. The mathematical relationship between the predictand and predictors is like given below:

$$Predictand_{(s)} = f\left(Predictor_{(s)}\right) \tag{4.1}$$

Similarly like the predictor, mathematically, if it is a single series then it is referred as uni-variate and multi-variate in case of several parallel series. Mathematical symbology for uni-variate and multi-variate predictands are represented using the following notations are given in the table 4.2.

Table 4.2: Type of predictands and notation

Predictand Type	Mathematical Symbol
Uni-variate Predictand	$\vec{y}$
Multi-variate Predictand	$\vec{Y}$

Equation (4.1) is a general mathematical relationship between predictands and predictors. Specific relationship between predictands and predictors are based on their uni-variate and multi-variate characteristics are like the equation discussed in subsequent sections.

$$\vec{y} = f(\vec{x}) \tag{4.2}$$

$$\vec{y} = f(\vec{X}) \tag{4.3}$$

$$\vec{Y} = f(\vec{X}) \tag{4.4}$$

Interconnection of local observations and large-scale circulation patterns is given by the equation 4.2, which is a simple linear statistical relationship [Von Storch et al., 1993, 1997]. Equation 4.3 represents the interconnection of local observations and large-scale circulation patterns. This equation represents the multiple linear regression equation. Equation 4.4 stands for a multivariate linear regression. In the equations 4.2 to 4.4, the predictands and predictors comprise of numerous observations and are represented by vectors (uni-variate) or matrices (multi-variate). For multivariate predictand, the matrix Y to refer to the time series of $\vec{y}(t)$, where columns of the matrix Y are vectors. Mathematical representation of matrix fields Y and X comprise of n measurements as given by following equations:

$$Y = [\vec{y}_1, \vec{y}_1, ...\vec{y}_n] \tag{4.5}$$

$$X = [\vec{x}_1, \vec{x}_1, ...\vec{x}_n] \tag{4.6}$$

In climate research, more than one climate variable (predictors) are used for the statistical downscaling. Downscaling uni-variate relationship of cimate variable (predictor) in downscaling process mostly produce less reliable climate projections, on the other hand the reliability of the multi-variate relationship of climate variable (predictors) in downscaling process generate more reliable climate projections.

4.4 Correlation

Correlation is the dependancy in form of a statistical relationship between two data-sets (random variable). Correlation analysis is statistical method, which is commonly used to determine the statistical relationship and the direction of the relationship between two data-sets. Hence, it is also referred as a bivariate correlation analysis in the statistics. Statistical relationship determined by the correlation analysis between two data-set is a linear statistical relation.

Correlations are useful because they can indicate a predictive relationship that can be used in practice. For example, precipitation in mountain region is based on the correlation between precipitation and humidity. A underlying relationship between humidity and precipitation in hilly regions causes high precipitation. Generally, the existence of a correlation is not sufficient to conclude the presence of a underlying relationship (i.e., correlation does not give the guarantee). Probabilistic independence between two data-set is essential mathematical property otherwise the data-sets show dependent mathematical nature, which make them unsuitable for the downscaling.

4.5 Correlation Coefficient

Strength (degree) of the correlation is represented by correlation coefficient. The value of the correlation coefficient is between +1 and -1. A value of $\pm$ 1 indicates a ideal degree of relationship between the two variables. As the correlation coefficient value tends towards zero, the relationship between the two variables becomes weaker. The direction of the relationship is designated by the sign of the coefficient; a positive sign refers a proportional relationship and a negative sign refers to inverse proportional relationship between the variables. Oftenly correlation coefficient is denoted by the mathematical notation r or ρ.

Usually, in statistics, correlation coefficient is measured by Pearson correlation coefficient, Kendall rank correlation coefficient and Spearman's rank correlation coefficient. Pearson correlation coefficient is the most common among them. Pearson correlation coefficient is responsive only for a linear relationship between two data-sets, while other two are more responsive sensitive to nonlinear relationships between data-sets

Correlation coefficient varies between +1 and -1. A verbally description of the strength of the correlation as the absolute value are given below:

Table 4.3: Correlation coefficient verbal classification

Verbally Description	**Value**
very weak	0.00 - 0.19
weak	0.20 - 0.39
moderate	0.40 - 0.59
strong	0.60 - 0.79
very strong	0.80 - 1.00

4.6 Pearson Correlation Coefficient

Pearson correlation coefficient is a parametric statistical test that is used to measure the degree of association between two variables (predictor and predictand) on the basis of their values (climate measurements). It is the most extensively used correlation statistic.

It measures the degree of the relationship between the two variables which are linearly related. For example, in the climate research, if one wants to measure how two variables like GHG concentration and a specified pollutant concentration like NO_x concentration are related to each other, Pearson cor-

relation coefficient can be used to estimate the degree of relationship between these two variables.

For the Pearson correlation coefficient, both variables must be normally distributed. Other assumptions include linearity and homoscedasticity. Linearity consider a straight line relationship between each of the two variables and homoscedasticity consider that data is equally distributed about the regression line.

The point-biserial correlation coefficient is used to estimate relationship when one of the variables is dichotomous. It is a specialized variant of Pearson correlation coefficient and used to assess the strength and direction of the association that exists between a dichotomous variable and a continuous variable. The following formula is used to calculate the Pearson r correlation:

$$r = \frac{\sum_{k=1}^{n}(x_k - \overline{x})(y_k - \overline{y})}{\sqrt{\sum_{k=1}^{n}(x_k - \overline{x})^2(y_k - \overline{y})^2}} \tag{4.7}$$

For the development of the computer program, Equation 4.7 can be expanded as:

$$r = \frac{n\sum_{k=1}^{n} x_k y_k - \sum_{k=1}^{n} x_k \sum_{k=1}^{n} y_k}{\sqrt{n\sum_{k=1}^{n} x_k^2 - (\sum_{k=1}^{n} x_k)^2}\sqrt{n\sum_{k=1}^{n} y_k^2 - (\sum_{k=1}^{n} y_k)^2}} \tag{4.8}$$

where: r = Pearson Correlation Coefficient

N = Number of observations

$\sum xy$ = summation of the products of paired measures

$\sum x$ = summation of x measures

$\sum y$ = summation of y measures

$\sum x^2$ = summation of squared x measures

$\sum y^2$ = summation of squared y measures

4.7 Spearman's Rank Correlation Coefficient

Spearman's rank correlation coefficient is a non-parametric test that is used to measure the degree of association between two variables on the basis of their ranks. Pearson correlation coefficient between set of ranked variables derived from the given set of variables on the basis of their value, is known as Spearman's rank correlation coefficient. Spearman's rank correlation coefficient is the base for nonparametric tests. Nonparametric tests refer to statistical testing methods in which it is not essential for the data to follow a normal distribution. The data in nonparametric tests are generally ordinal in nature that means that data do not rely on numbers, but rather on their ranking in dataset.

Set of variables are X and Y, representing predictors and predictands respectively. Their corresponding ranked variable sets are represented below.

$$Ranking(X) = \mathbb{X}$$
$$Ranking(Y) = \mathbb{Y}$$

Spearman's Rank Coefficient is determined using following equation:

$$r_s = \rho_{\mathbb{X},\mathbb{Y}} = \frac{Cov(\mathbb{X},\mathbb{Y})}{\sigma_{\mathbb{X}}\sigma_{\mathbb{Y}}} \tag{4.9}$$

where: r_s = Spearman's Rank Correlation Coefficient

$\rho_{\mathbb{X},\mathbb{Y}}$ = Pearson Correlation Coefficient of ranked variables.

$Cov(\mathbb{X},\mathbb{Y})$ = Covariance of the rank variables.

$\sigma_{\mathbb{X}}$ and $\sigma_{\mathbb{Y}}$ = Standard deviations of the ranked variables.

If the ranked values of the two variables for a set of n measurements are x_k and y_k, with $d_k = y_k - x_k$, then the Spearman's rank correlation coefficient is

defined as

$$r_s = \rho = 1 - \frac{6\sum_{k=1}^{n} d_k^2}{n(n^2-1)} \tag{4.10}$$

where: r_s = Spearman's Rank Correlation Coefficient

d_k = Difference between the ranks of corresponding variables

n = Number of Observations

Where notation r_s denotes Spearman's rank correlation coefficient between the rankings $\vec{x}$ and $\vec{y}$. The assumptions of the Spearman's rank correlation coefficient are that the measurements on one variable must be monotonically related to the other variable and data must be ordinal.

For this research work, guidelines given by Yu et al. [2017] have been used to investigate the degree of the relationship between GCM outputs using Spearman's rank correlation coefficient. The value of correlation coefficient between 0 and 0.40 represent a low linear association, coefficients between 0.40 and 0.70 represent a significant association, and value of correlation coefficient between 0.7 and above represent a large association or relationship.

4.8 Multiple Linear Regression

4.8.1 Mathematical Model

Multiple linear regression (MLR) is a generalized extension of simple linear regression to the case in which there are various predictors for the development of the model. For example, for two predictor variables X_1 and X_2, and one predictand y, the regression model can be expressed as:

$$y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + e \tag{4.11}$$

Above equation given for example comprises of a deterministic component involving the three regression coefficients (α_0, α_1 and α_2) and ϵ involving the residual terms.

As equation (4.11) is a representation by bi-variate linear regression model, likewise the general equation for the multiple linear regression (MLR) model is represented like equation (4.12) is presented below:

$$y = \alpha_0 + \sum_{k=1}^{n} \alpha_k x_k + \epsilon \tag{4.12}$$

where: y = Predictand (climate variable)

x_k = k^{th} Predictor

α_k = k^{th} Regression coefficient

n = Number of predictors

ϵ = Residual terms

The matrix representation of the multiple linear regression model with resemblance of the (4.12) is represented like:

$$y = \underset{(1\times(n+1))}{\boldsymbol{X}} \underset{((n+1)\times 1)}{\boldsymbol{\alpha}} + \epsilon \tag{4.13}$$

$$y = \begin{bmatrix} 1 & x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} + \epsilon \tag{4.14}$$

The predictor variables can be either continuous or categorical. In the case of the categorical type, these variables are required to be encoded as dummy variables. The dependent variable y must be measured on a continuous scale. The residual terms represent the difference between the predicted and observed values (predictors) of measurements (X_k). These terms are assumed to be independently and identically distributed with zero mean and variance, and account for natural variability as well as maybe measurement error.

4.8.2 Model Assumptions

Following are the assumptions for multiple linear regression

1. There must be a linear relationship between predictand and predictors. Scatterplots can be helpful for revealing the linear/curvilinear relationship.

2. The residuals are normally distributed.

3. The independent variables should not be highly correlated with each other. Henceforth, independent variables do not exhibit multicollinearity.

4. Variances of error terms are similar across the values of the independent variables. This property is termed as homoscedasticity. A plot of standardized residuals versus predicted values can reveal that whether points are equally distributed across all values of the independent variables.

4.9 Multivariate Linear Regression

Multivariate linear regression (MvLR) refers to the modeling of data wherein an outcome is measured for the same phenomena at multiple times (repeated measurements), or the modeling of nested/clustered data, wherein there are multiple group of measurements in each cluster. A multivariate linear regression model is generalized as:

$$\boldsymbol{Y} = \boldsymbol{\alpha}\boldsymbol{X} + \boldsymbol{\epsilon} \tag{4.15}$$

$$\boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_k \end{bmatrix}_{(k\times 1)} \tag{4.15a}$$

$$\boldsymbol{\alpha} = \begin{bmatrix} \boldsymbol{\alpha}_0 & \boldsymbol{\alpha}_1 & \boldsymbol{\alpha}_2 & \cdots & \boldsymbol{\alpha}_n \end{bmatrix}_{(1\times(n+1))} \tag{4.15b}$$

$$\boldsymbol{X} = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}_{((n+1)\times 1)} \tag{4.15c}$$

$$\boldsymbol{\epsilon} = \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \boldsymbol{\epsilon}_3 \\ \vdots \\ \boldsymbol{\epsilon}_p \end{bmatrix}_{(p\times 1)} \tag{4.15d}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_p \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha}_0 & \boldsymbol{\alpha}_1 & \boldsymbol{\alpha}_2 & \cdots & \boldsymbol{\alpha}_n \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \boldsymbol{\epsilon}_3 \\ \vdots \\ \boldsymbol{\epsilon}_p \end{bmatrix} \tag{4.16}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_p \end{bmatrix} = \boldsymbol{\alpha}_0 x_0 + \boldsymbol{\alpha}_1 x_1 + \boldsymbol{\alpha}_2 x_2 + \cdots + \boldsymbol{\alpha}_n x_n + \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \boldsymbol{\epsilon}_3 \\ \vdots \\ \boldsymbol{\epsilon}_p \end{bmatrix} \tag{4.17}$$

$$\boldsymbol{Y} = \boldsymbol{\alpha}_0 x_0 + \boldsymbol{\alpha}_1 x_1 + \boldsymbol{\alpha}_2 x_2 + \ldots + \boldsymbol{\alpha}_n x_n + \boldsymbol{\epsilon} \tag{4.18}$$

$$\boldsymbol{Y} = \sum_{k=0}^{n} \boldsymbol{\alpha}_k x_k + \boldsymbol{\epsilon} \tag{4.19}$$

where: $\boldsymbol{Y}$ = Predictand set with k predictands : $(y_1, y_2, \ldots, y_p)$

x_k = k^{th} Predictor

$\boldsymbol{\alpha}_k$ = k^{th} Regression coefficient set with k elements

$\boldsymbol{\epsilon}$ = Residual set with p elements

n = Number of predictors

Equation (4.19) is a set of p elements. Each element has its own model equation, which is a multiple linear regression model.

$$\boldsymbol{\alpha} = \begin{bmatrix} \boldsymbol{\alpha}_0 & \boldsymbol{\alpha}_1 & \boldsymbol{\alpha}_2 & \cdots & \boldsymbol{\alpha}_n \end{bmatrix} = \begin{bmatrix} \alpha_{10} & \alpha_{11} & \alpha_{12} & \ldots & \alpha_{1n} \\ \alpha_{20} & \alpha_{21} & \alpha_{22} & \ldots & \alpha_{2n} \\ \alpha_{30} & \alpha_{31} & \alpha_{32} & \ldots & \alpha_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_{p0} & \alpha_{p1} & \alpha_{p2} & \ldots & \alpha_{pn} \end{bmatrix} \tag{4.20}$$

As expressed in equation (4.15b), $\boldsymbol{\alpha}$ is set of $(n+1)$ elements which are itself set of p elements. This is expressed in the equation (4.20).

By the combining the equation (4.16) and the equation (4.20), following equation can be obtained:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_p \end{bmatrix} = \begin{bmatrix} \alpha_{10} & \alpha_{11} & \alpha_{12} & \dots & \alpha_{1n} \\ \alpha_{20} & \alpha_{21} & \alpha_{22} & \dots & \alpha_{2n} \\ \alpha_{30} & \alpha_{31} & \alpha_{32} & \dots & \alpha_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_{p0} & \alpha_{p1} & \alpha_{p2} & \dots & \alpha_{pn} \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \vdots \\ \epsilon_p \end{bmatrix} \tag{4.21}$$

By solving the equation (4.21), the set of solution is obtained as:

$$\begin{aligned} y_1 &= \alpha_{10}x_0 + \alpha_{11}x_1 + \alpha_{12}x_2 + \cdots + \alpha_{1n}x_n + \epsilon_1 \\ y_2 &= \alpha_{20}x_0 + \alpha_{21}x_1 + \alpha_{22}x_2 + \cdots + \alpha_{2n}x_n + \epsilon_2 \\ y_3 &= \alpha_{30}x_0 + \alpha_{31}x_1 + \alpha_{32}x_2 + \cdots + \alpha_{3n}x_n + \epsilon_3 \\ \vdots & \qquad\qquad\quad \vdots \\ y_p &= \alpha_{p0}x_0 + \alpha_{p1}x_1 + \alpha_{p2}x_2 + \cdots + \alpha_{pn}x_n + \epsilon_p \end{aligned} \tag{4.22}$$

$$y_j = \alpha_{j0}x_1 + \alpha_{j1}x_1 + \alpha_{j2}x_2 + \cdots + \alpha_{jn}x_n + \epsilon_j \tag{4.23}$$

$$y_j = \sum_{k=0}^{n} \alpha_{jk}x_k + \epsilon_j \tag{4.24}$$

Above presented equation (4.21) and equation (4.22) are the matrix representation of multi-variate linear regression (MvLR) model, which are suitable for their implementation using computer program.

Equation (4.23) and equation (4.24) are referring that MvLR model can be obtained by applying MLR model solution on predictand set ($\boldsymbol{Y}$) with p elements $(y_1, y_2, \cdots, y_p)$.

4.10 Model Solution

General multiple linear regression model is represented as equation (4.12). there are n predictors, $(n+1)$ regression coefficients and one residual term.

MLR models are often used as empirical models or approximation functions. This is the reason that the true functional relationship between y and $x_1, x_2, \ldots, x_n$ is unknown.

Complex models may often be analyzed by MLR technique like in equation by binomial model and in equation by quadratic polynomial model.

$$y = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \epsilon \tag{4.25}$$

$$y = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3 + \alpha_4 x^4 + \epsilon \tag{4.26}$$

Equation (4.25) can be rewritten like

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \epsilon \tag{4.27}$$

By taking the assumption $x_1 = x$ and $x_2 = x^2$.

And equation (4.26) can be rewritten like

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4 + \epsilon \tag{4.28}$$

By taking the assumption $x_1 = x$, $x_2 = x^2$, $x_3 = x^3$ and $x_4 = x^4$.

Models that have interaction between predictors can also be analysed by MLR methods. For example models that are like model represented in the equation below:

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_{12} x_1 x_2 + \epsilon \tag{4.29}$$

This model given in equation (4.29) can be rewritten like

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \epsilon \tag{4.30}$$

By taking the assumption $x_3 = x_1 x_2$ and $\alpha_3 = \alpha_{12}$.

Which is finally a linear regression model equation. Same thing can also be achieved with higher polynomial order models, multi-interaction models and combination of higher polynomial order and multi-interaction models. In end a final multiple linear regression model equation is obtained.

4.10.1 Estimation of the model parameters by least-squares method

Regression coefficients can be estimated by least-square method. If there are m number of observations and n number of predictors (regressors), then

$$m > n \tag{4.31}$$

$$\boldsymbol{y} = \{y_1, y_2, \ldots, y_m\} \tag{4.32}$$

$$\boldsymbol{x}_i = \{x_{i1}, x_{i2}, \ldots, x_{im}\} \tag{4.33}$$

and also $\boldsymbol{\epsilon}$ in the model has uncorrelated errors and

$$E(\boldsymbol{\epsilon}) = 0 \tag{4.34}$$

$$Var(\boldsymbol{\epsilon}) = \rho^2 \tag{4.35}$$

As regression data come from the climate observation study, most of the predictors (regressos) will be random variables. So observations of each predictors will be independent and regression coefficients ($\boldsymbol{\alpha}$) or variance (σ^2) will not affect the distribution.

By assuming that the mean of $(\boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_1\boldsymbol{x}_1 + \cdots + \boldsymbol{\alpha}_n\boldsymbol{x}_n)$ and variance (σ^2) be normal with predictors $(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$ imply that predictand $(\boldsymbol{y})$ is conditionally distributed. However, these variables are set of observations and contain m obervations. So by considering the simple multiple linear regression model equation for each observation are like:

$$\begin{aligned} y_j &= \alpha_0 + \alpha_1 x_{j1} + \alpha_2 x_{j2} + \cdots + \alpha_n x_{jn} + \epsilon_j \\ &= \alpha_0 + \sum_{k=1}^{n} \alpha_k x_{jk} + \epsilon_j \end{aligned} \tag{4.36}$$

where: $\quad j = 1, 2, \ldots, m$

$$\epsilon_j = y_j - \alpha_0 - \sum_{k=1}^{n} \alpha_k x_{jk} \tag{4.37}$$

The least-square function is formulated like given below in equation :

$$\boldsymbol{S}\big(\alpha_0, \alpha_1, \ldots, \alpha_n\big) = \sum_{j=1}^{m} \epsilon_j^2 \tag{4.38}$$

By applying equation (4.37) to equation (4.38) the new derived equation is :

$$\boldsymbol{S}\big(\alpha_0, \alpha_1, \ldots, \alpha_n\big) = \sum_{j=1}^{m} \left(y_j - \alpha_0 - \sum_{k=1}^{n} \alpha_k x_{jk} \right)^2 \tag{4.39}$$

$\boldsymbol{S}$ is a minimization function with respect to $(\alpha_0, \alpha_1, \ldots, \alpha_n)$. As for the minimization:

$$\left. \frac{d_{\boldsymbol{S}}}{d_{\alpha_0}} \right|_{\hat{\alpha_0}, \hat{\alpha_1}, \ldots, \hat{\alpha_n}} = 0 \tag{4.40}$$

By applying condition of equation (4.40) to the equation (4.39) a new derived equations are:

$$\frac{d\boldsymbol{S}}{d_{\alpha_0}}\bigg|_{\widehat{\alpha_0},\widehat{\alpha_1},\ldots,\widehat{\alpha_n}} = 2\sum_{j=1}^{m}\left(y_j - \widehat{\alpha_0} - \sum_{k=1}^{n}\widehat{\alpha_k}x_{jk}\right) = 0 \tag{4.41}$$

and

$$\frac{d\boldsymbol{S}}{d_{\alpha_j}}\bigg|_{\widehat{\alpha_0},\widehat{\alpha_1},\ldots,\widehat{\alpha_n}} = 2\sum_{j=1}^{m}\left(y_j - \widehat{\alpha_0} - \sum_{k=1}^{n}\widehat{\alpha_k}x_{jk}\right)x_{jk} = 0 \tag{4.42}$$

Solution obtained by solving equation (4.41) and equation (4.42) like equation set (4.43) described next.

Equation (4.43) comprises of $n+1$ sub-equations one for each regression coefficients. The solutions for sub-equations will be the least square estimators for $\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_n$. For the computation purpose, the matrix notation of the multiple linear regression model is more suitable.

$$
\begin{array}{ccccccccccc}
\sum_{j=1}^{m} y_j & = & n\widehat{\alpha_0} & + & \widehat{\alpha_1}\sum_{j=1}^{m} x_{j1} & + & \widehat{\alpha_2}\sum_{j=1}^{m} x_{j2} & + & \cdots & + & \widehat{\alpha_n}\sum_{j=1}^{m} x_{jn} \\
\sum_{j=1}^{m} x_{j1}y_j & = & \widehat{\alpha_0}\sum_{j=1}^{m} x_{j1} & + & \widehat{\alpha_1}\sum_{j=1}^{m} x_{j1}^2 & + & \widehat{\alpha_2}\sum_{j=1}^{m} x_{j1}x_{j2} & + & \cdots & + & \widehat{\alpha_n}\sum_{j=1}^{m} x_{j1}x_{jn} \\
\sum_{j=1}^{m} x_{j2}y_j & = & \widehat{\alpha_0}\sum_{j=1}^{m} x_{j2} & + & \widehat{\alpha_1}\sum_{j=1}^{m} x_{j2}x_{j1} & + & \widehat{\alpha_2}\sum_{j=1}^{m} x_{j2}^2 & + & \cdots & + & \widehat{\alpha_n}\sum_{j=1}^{m} x_{j2}x_{jn} \\
\vdots & = & \vdots & & \vdots & & \vdots & & \ddots & & \vdots \\
\sum_{j=1}^{m} x_{jn}y_j & = & \widehat{\alpha_0}\sum_{j=1}^{m} x_{jn} & + & \widehat{\alpha_1}\sum_{j=1}^{m} x_{jn}x_{j1} & + & \widehat{\alpha_2}\sum_{j=1}^{m} x_{jn}x_{j2} & + & \cdots & + & \widehat{\alpha_n}\sum_{j=1}^{m} x_{jn}^2
\end{array} \tag{4.43}
$$

Matrix notation of the multiple linear regression model in context of data (observation) is represented like equation given below:

$$\underset{(m\times 1)}{\boldsymbol{y}} = \underset{(m\times(n+1))}{\boldsymbol{X}} \quad \underset{((n+1)\times 1)}{\boldsymbol{\alpha}} + \underset{(m\times 1)}{\boldsymbol{\epsilon}} \tag{4.44}$$

The element of equation (4.44) are represented using the following matrices:

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y \end{bmatrix}_{m\times 1} \tag{4.44a}$$

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1n} \\ 1 & x_{21} & x_{22} & \dots & x_{2n} \\ 1 & x_{31} & x_{32} & \dots & x_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix}_{m\times(n+1)} \tag{4.44b}$$

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}_{(n+1)\times 1} \tag{4.44c}$$

$$\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \vdots \\ \epsilon_m \end{bmatrix}_{m\times 1} \tag{4.44d}$$

In the equations set (4.44), predictand ($\boldsymbol{y}$) has m vectors as climate observations, predictor set ($\boldsymbol{X}$) has $n+1$ elements as predictors $(x_0, x_1, x_2, \ldots, x_n)$, where $x_0 = \boldsymbol{I}_{m\times 1}$ and rest of predictors have m vectors as climate observations. Regression coefficient $\boldsymbol{\alpha}$ is a matrix with dimensions $(n+1)\times 1$ and has $(n+1)$ vector of regression coefficients. Residual ($\boldsymbol{\epsilon}$) is a matrix with dimensions $(m \times 1)$ and has (m) vector elements as residual terms.

For finding the minima of the vector of least square estimator $\widehat{\boldsymbol{\alpha}}$ the solution is like given below:

$$\boldsymbol{S}(\boldsymbol{\alpha}) = \sum_{j=1}^{m} \epsilon_j^2 \tag{4.45}$$

As

$$\sum_{j=1}^{m} \epsilon_j^2 = \boldsymbol{\epsilon}'\boldsymbol{\epsilon} \tag{4.46}$$

where:

$$\boldsymbol{\epsilon}' = Transpose(\boldsymbol{\epsilon}) \tag{4.47}$$

and

$$\boldsymbol{\epsilon} = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha} \tag{4.48}$$

By applying equations (4.46), equation (4.47) and (4.48) to the equation (4.45):

$$\boldsymbol{S}(\boldsymbol{\alpha}) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}) \tag{4.49}$$

$$\boldsymbol{S}(\boldsymbol{\alpha}) = \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{\alpha}'\boldsymbol{X}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{X}\boldsymbol{\alpha} + \boldsymbol{\alpha}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\alpha} \tag{4.50}$$

$$(\boldsymbol{\alpha}'\boldsymbol{X}'\boldsymbol{y})' = \boldsymbol{y}'\boldsymbol{X}\boldsymbol{\alpha} \tag{4.51}$$

Equation (4.51) depicts algebraic property $(M_1M_2)' = M_2'M_1'$ of matrix operations, and $\boldsymbol{\alpha}'\boldsymbol{X}'\boldsymbol{y}$ is a 1×1 matrix or a scalar, equations (4.49) - (4.51)

can be represented using the following equations:

$$\boldsymbol{\alpha}'\boldsymbol{X}'\boldsymbol{y} = (\boldsymbol{\alpha}'\boldsymbol{X}'\boldsymbol{y})' = \boldsymbol{y}'\boldsymbol{X}\boldsymbol{\alpha} \tag{4.52}$$

So

$$\boldsymbol{S}(\boldsymbol{\alpha}) = \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{\alpha}'\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{\alpha}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\alpha} \tag{4.53}$$

As for minimization of $\boldsymbol{S}(\boldsymbol{\alpha})$ must satisfy the following condition

$$\left.\frac{d\boldsymbol{S}}{d\boldsymbol{\alpha}}\right|_{\widehat{\boldsymbol{\alpha}}} = 0 \tag{4.54}$$

By applying condition of equation (4.54) on equation (4.53):

$$\left.\frac{d\boldsymbol{S}}{d\boldsymbol{\alpha}}\right|_{\widehat{\boldsymbol{\alpha}}} = -2\boldsymbol{X}'\boldsymbol{y} + 2\boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\alpha}} = 0 \tag{4.55}$$

Which can be simplified as:

$$\boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\alpha}} = \boldsymbol{X}'\boldsymbol{y} \tag{4.56}$$

$$\boxed{\widehat{\boldsymbol{\alpha}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}} \tag{4.57}$$

$$\boldsymbol{X}' = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & x_{31} & \dots & x_{m1} \\ x_{12} & x_{22} & x_{32} & \dots & x_{m2} \\ \vdots & \vdots & \vdots & \ddots & \dots \\ x_{1n} & x_{2n} & x_{3n} & \dots & x_{mn} \end{bmatrix}_{(n+1)\times m} \tag{4.57a}$$

$$\boldsymbol{X}'\boldsymbol{X} = \begin{bmatrix} m & \sum_{j=1}^{m} x_{j1} & \sum_{j=1}^{m} x_{j2} & \dots & \sum_{j=1}^{m} x_{jn} \\ \sum_{j=1}^{m} x_{j1} & \sum_{j=1}^{m} x_{j1}^2 & \sum_{j=1}^{m} x_{j1}x_{j2} & \dots & \sum_{j=1}^{m} x_{j1}x_{jn} \\ \sum_{j=1}^{m} x_{j2} & \sum_{j=1}^{m} x_{j2}x_{j1} & \sum_{j=1}^{m} x_{j2}^2 & \dots & \sum_{j=1}^{m} x_{j2}x_{jn} \\ \vdots & \vdots & \vdots & \ddots & \dots \\ \sum_{j=1}^{m} x_{jn} & \sum_{j=1}^{m} x_{jn}x_{j1} & \sum_{j=1}^{m} x_{jn}x_{j2} & \dots & \sum_{j=1}^{m} x_{jn}^2 \end{bmatrix}_{(n+1)\times(n+1)} \tag{4.57b}$$

Equation (4.57b) represent that $\boldsymbol{X}'\boldsymbol{X}$ ia a $(n+1) \times (n+1)$ symmetric matrix.

The fitted regression model corresponding to the matrix form of the predictand $\boldsymbol{y}$ and the predictors $\boldsymbol{X}$:

$$\widehat{\boldsymbol{y}} = \boldsymbol{X}\widehat{\boldsymbol{\alpha}} \tag{4.58}$$

By applying equation (4.57) to the equation (4.58).

$$\widehat{\boldsymbol{y}} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \tag{4.59}$$

$$\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' \tag{4.60}$$

$\boldsymbol{H}$ is known as hat matrix with dimension $m \times m$.

$$\widehat{\boldsymbol{y}} = \boldsymbol{H}\boldsymbol{y} \tag{4.61}$$

Now for determining the residual term: Now for determining the residual terms, equation (4.63) is derived as follows:

$$\boldsymbol{\epsilon} = \boldsymbol{y} - \widehat{\boldsymbol{y}} \tag{4.62}$$

$$\boldsymbol{\epsilon} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\alpha}} \tag{4.62a}$$

$$\boldsymbol{\epsilon} = \boldsymbol{y} - \boldsymbol{H}\boldsymbol{y} \tag{4.62b}$$

$$\boldsymbol{\epsilon} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{y} \tag{4.62c}$$

$$\boxed{\boldsymbol{\epsilon} = \big(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\big)\boldsymbol{y}} \tag{4.63}$$

By the data matrix shown in equation (4.44a), equation (4.44b), equation (4.57a) and equation (4.57b) computation of $\boldsymbol{\epsilon}$ done. These equations are very computational friendly for the implementation purpose.

Aforementioned mathematical solution is for the MLR model. As the generalization of MLR model, MvLR model refers the modeling of nested/clustered data, wherein multiple group of measurements are taken. If there are p group of measurements are present, then above solution is applied for each group.

Mathematical solution discussed till now is for the MLR model. MvLR model refers the modeling of nested/clustered data , wherein there are multiple group of measurements are taken. If there are p group of measurements are present, then above solution is applied for each group.

4.11 Concluding Remarks

In this chapter, underlying mathematical concepts of the statistical downscaling approach adopted by EMSDM have been discussed. Suitable Preditors (GCM output parameters) are selected on the basis of Correlation Statistics. These statistics are the extensively used for selecting suitable predictors from the available GCM outputs (predictors) Chen et al. [2012]; Meenu et al. [2013]. Multiple and multivariate regression techniques are widely adopted by Statistical Downscaling Approaches for model development [Tang et al., 2016a; Yang et al., 2017a].

In the subsequent chapter 5, the basic framework of Efficient Multi-site Statistical Downscaling Model (EMSDM) is discussed in detail.

Chapter 5

Efficient Multi-site Statistical Downscaling Model (EMSDM)

5.1 General

In this chapter, proposed computational framework namely Efficient Multi-site Statistical Downscaling Model (EMSDM) has been discussed. EMSDM employs automation to carry out downscaling of multiple grids. Proposed framework can easily incorporate different GCMs for downscaling and different types of dataset for downscaling. In the subsequent section, underlying steps of the framework and their flow have been discussed. Detailed implementation of the framework is discussed in chapter 6.

5.2 Framework

Efficient Multi-site Statistical Downscaling Model is depicted in Figure 5.1. Owing to wide spread usage, requirement of nominal computational resources and possibility of extraction of site-specification information, it adopts the statistical downscaling approach for high resolution downscaling of given AOI. As conspicuous from Figure 5.1, some of steps of the EMSDM are automated to accomplish downscaling for all the grids or locations of a given AOI. Steps to carry out downscaling using EMSDM are discussed in subsequent sub-sections.

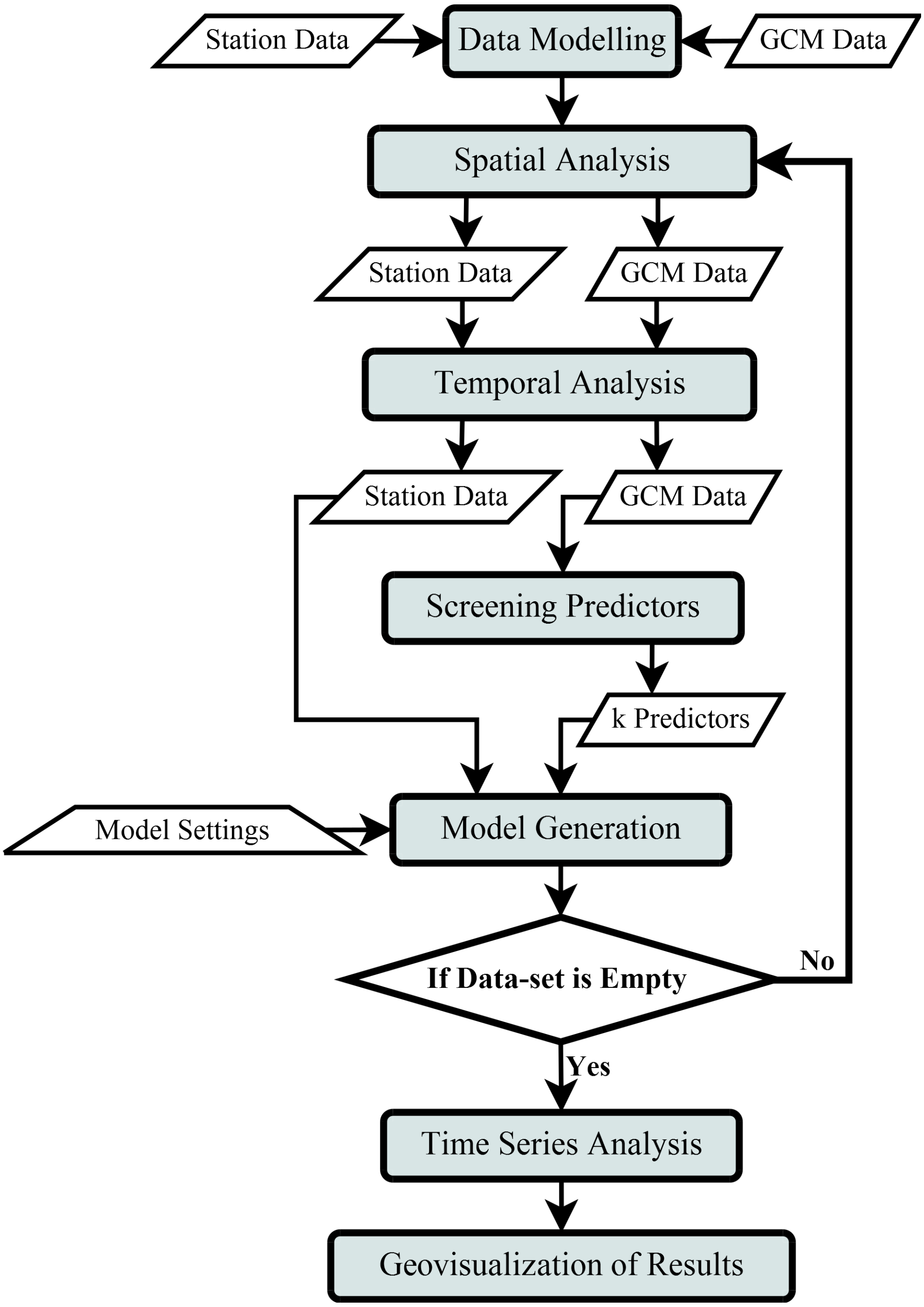


Figure 5.1: Framework for EMSDM

5.2.1 Data Modelling

GCM and climatological data is available is different formats that are spatially referenced in grid form. However these dataset are not adequately structured for downscaling for AOI. In data modelling step, available data are structured suitably for the framework. Moreover, metadata of dataset used for downscaling is not available. Structuring of dataset in suitable spatial format along with its metadata is automated through development and implementation of computer programmes.

5.2.2 Spatial Analysis

In this process, spatial metadata prepared using the previous step is used to validate and spatially map the GCM dataset with local climate data like IMD data through GIS based operations like overlay, encoding of local climate dataset with GCM dataset etc. Spatial analysis is important for automating the mapping the relevant GCM dataset with the corresponding local climate dataset of interest. This process is useful in automating the time-intensive manual validation and mapping of GCM dataset with local dataset for large areas.

5.2.3 Temporal Analysis

GCM data and local climate data are available at different time scale viz. years. In this step, temporal mapping of the GCM dataset with local climate data like IMD data is carried out. Moreover statistical aggregation of daily climate dataset is automated to develop the monthly, yearly and decadal climate dataset. Through this process monthly, yearly and decadal climate dataset is readily available for further applications like trend visualization, downscaling etc. Morover, this process is also helpful in significantly saving the computational time required for preparation of data-sets.

5.2.4 Screening Predictors

GCM datasets comprises of various climate parameters. For example, CanESM2 data-set comprises of 26 climate parameters. In context of downscaling, these climate parameters are known as predictors. Moreover effect of these predictors varies geographically. Henceforth, the suitable set of predictors vary from grid to grid.

In this step, non-dominating predictors are screened out from the set of predictors so as to obtain set of dominating predictors for downscaling process. The major criterion of selecting dominating predictors is that the dominating predictors are strongly correlated with the predictands [Grimes et al., 2003; Yang et al., 2017a].

For the proposed framework, screening of predictors is carried out by two type of statistics.

1. Parametric Statistics
2. Non-parametric Statistics

Generally selection of the type of statistics for determining the correlation coefficients requires manual intervention of the user. Extreme events acts as the outliers. One outlier can affect the parametric statistics by inflating the variance and enhance the error term. This can invalidate the conclusion drawn by the parametric statistics. Henceforth in case of large occurrence of extreme events, non-parametric statistics is more suitable for the selection of the predictors. For parametric statistics, Pearson Correlation Coefficient and for non-parametric statistics, Spearman's Correlation Coefficient are implemented by the model to estimate the correlation coefficients, which are subsequently used to screen out non-relevant predictors and selecting the dominating predictors.

Threshold value for the selection of the set of predictors is automatically determined by an algorithm. Threshold value adopted by the algorithm is

based on the fact that set of selected predictors should not be the null set. Generally, the threshold value of correlation coefficient for screening set of predictors, has value more than 0.7. Moreover, factor analysis is used by the model for the dimensionality reduction of selected predictor set in the case when cardinality of the set of predictors generated from the aforementioned algorithm is high (for example larger than 12).

5.2.5 Model Generation

In this step, downscaling model is developed using Multiple Linear Regression (MLR) and Multivariate Linear Regression Analysis (MvLR). Screened predictors and daily local station dataset (predictants) are used to develop the model. EMSDM can be used to develop the downscaling model using screened predictors and daily local climate dataset (predictants) as well as aggregated local climate dataset (predictants) like monthly data-sets, decadal data-sets etc. In this step, computationally intensive mathematical operations are applied on large predictand and predictors data-sets inputted as matrices of large dimensions. These complex mathematical operations are implemented in 'C' programming language.

5.2.6 Time Series Analysis

In this step, model developed in previous step is utilized to generate time series of the climatological parameters for future years,decades for the specified RCP scenario. Model is fed with the specified RCP scenario data-set and model coefficients to generate the time series of the climatological parameters for future.

5.2.7 Geovisualiazation of Results

Gevisualization of results of downscaling is carried out using Web GIS application. The Web-GIS application is developed using open-source python based web-framework namely Django. User can query the outputs available from the developed Web GIS application.

5.3 Concluding Remarks

In this chapter, underlying steps of EMSDM has been discussed. Proposed framework is able to carry out multi-site downscaling for given area of interest (AOI). Different processes of EMSDM are automated to carryout downscaling, henceforth manual intervention is not required to develop downscaling model.

EMSDM is implemented in “C” programming language. Generally, the climate studies require handling of large amount of data and intensive processing. Application developed in “C” programming language supports fast processing and can efficiently handles large quantum of GCM and local data-sets.

Chapter 6

Implementation

6.1 General

In this chapter, implementation of EMSDM is described in form for primary algorithms as pseudo codes. Local precipitation dataset (predictand) of India acquired from India Meteorological Department (IMD) and GCM dataset (predictors) namely Second generation Canadian Earth System Model (CanESM2) acquired from Canadian Centre for Climate Modelling and Analysis (CCCma) have been taken for validating the EMSDM.

As a primary climate data, precipitation data available from IMD is structured as the grids with grid resolution $0.5° \times 0.5°$. Each grid may contain one or more observation stations. Temporal resolution of climate data-set obtained from IMD is taken as one day. Temporal duration of the climate data-set is from year 1971 to 2005.

As a GCM, CanESM2 is obtained from CCCma is also structured as the grids with grid resolution $2.7906° \times 2.8125°$. Temporal resolution of predictors data-set obtained from this GCM is taken as one day.

Well-known text (WKT) as a text markup language for creating spatial data used in this research work. Standardized formats of WKT are discussed in next section which are subsequently used for generation spatial grid data.

6.2 Well-known text (WKT)

WKT is a text markup language. It is used for spatially representing, referencing and transforming the spatial data. Well-known binary (WKB) is the binary equivalent of WKT and is used to transfer and store the same information on databases. WKT and WKB formats were formerly defined by the Open Geospatial Consortium (OGC). OGC provide the specification this format in their Coordinate Transformation Service and Simple Feature Access specifications. However, presently these formats are standardized through ISO/IEC 13249-3:2016 and ISO 19162:2015 standards. WKT can represent the following types of geometric objects like (ISO/IEC 13249-3: 2016):

1. Geometry
2. Point
3. LineString
4. Polygon
5. MultiPoint
6. MultiLineString
7. MultiPolygon
8. GeometryCollection
9. Circular String
10. CompoundCurve
11. CurvePolygon
12. MultiCurve
13. MultiSurface
14. Curve
15. Surface
16. PolyhedralSurface
17. TIN
18. Triangle

Table 6.1: Basic Geometry Types of WKT Format

Geometry primitives (2D)	
Type	**Examples**
Point	POINT(30 10)
LineString	LINESTRING (30 10, 10 30, 40 40)
Polygon	POLYGON ((30 10, 40 40, 20 40, 10 20, 30 10))
	POLYGON ((35 10, 45 45, 15 40, 10 20, 35 10), (20 30, 35 35, 30 20, 20 30))

Table 6.2: Complex Geometry Types of WKT Format

Geometry primitives (2D)	
Type	**Examples**
MultiPoint	MULTIPOINT ((10 40), (40 30), (20 20), (30 10))
	MULTIPOINT (10 40, 40 30, 20 20, 30 10)
MultiLineString	MULTILINESTRING ((10 10, 20 20, 10 40), (40 40, 30 30, 40 20, 30 10))
MultiPolygon	MULTIPOLYGON (((30 20, 45 40, 10 40, 30 20)), ((15 5, 40 10, 10 20, 5 10, 15 5)))
	MULTIPOLYGON (((40 40, 20 45, 45 30, 40 40)), ((20 35, 10 30, 10 10, 30 5, 45 20, 20 35), (30 20, 20 15, 20 25, 30 20)))

6.3 Pre-Processing of Local Data-sets (Predictand)

For the present research work, data-set as a predictand is acquired from IMD. Data-set is in form of 35 files, each corresponding to precipitation data measurement from year 1971 to year 2005. Each of these 35 files comprises of daily observation value for precipitation in binary format. These observational measurement values are assigned in form of data-matrix with dimension (65×69), where each data-matrix element is representation of a grid. Henceforth, there are 65×69 rectangular grids comprise of the precipitation data of India and its vicinity. Values associated with grids are of following types:

1. For Valid $value > 0$.

2. For Invalid $value = -999 \Rightarrow NULL$.

Total numbers of days from year 1971 to 2005 is equal to 12874. For each day, number of data points for precipitation are 4485 $(= 65 \times 69)$.

A procedure to extract information of valid grids from the raw IMD data has been developed. This procedure is given in Pseudo Code 1. By applying this procedure, one can obtain the maximum number of grids which has the valid value and within the administrative boundary of India. For maintaining the generality of the pseudo codes, the constants **Row** and **Col** of the pseudo codes used for grid-matrix is not fixed to the value of grid-matrix dimensions IMD's precipitation data, which are 65 for *row* and 69 for *col*.

However some of stations were established by IMD in later years. So number of these grids which are established in later years can be determined using procedure given in Pseudo Code 2. From the algorithms discussed in Pseudo Codes 1 and 2, the information regarding the minimum and maximum number of the grids that are established in later years and that corresponds to India have been acquired. In this way, the framework is flexible as well as scalable to handle the varying cardinality of the datasets.

Pseudo Code 1. To extract valid grids from the raw IMD data

```
SET Count → 0
SET Grid_Count → 0
SET File_List → [IMD Files]
foreach file in File_List do
    foreach DataSet[ROW][COL] in file do
        for row = 1 to ROW do
            for col = 1 to COL do
                if DataSet[row][col] ≥ 0 then
                    Count = Count + 1
                end
            end
        end
        if Grid_Count < 0 then
            Grid_Count = Count
        end
        Count = 0
    end
end
PRINT Grid_Count
```

Pseudo Code 2. To find number of grids which are established in later years

```
SET Count → 0
SET Grid_Count_Min → 0
SET File_List → [IMD Files]
foreach file in File_List do
    foreach DataSet[ROW][COL] in file do
        for row = 1 to ROW do
            for col = 1 to COL do
                if DataSet[row][col] ≥ 0 then
                    Count = Count + 1
                end
            end
        end
        if Grid_Count < Grid_Count_Min then
            Grid_Count_Min = Count
        end
        Count = 0
    end
end
New_Grids = Grid_Count - Grid_Count_Min
PRINT New_Grids
```

Procedure to extract information of number of valid grids from the raw IMD data has been developed. This procedure is given in Pseudo Code 3. By applying this procedure, $Data\_Set\_Number$ named global variable subsequently set to the value of number of valid grids in the IMD data-set, which specifically lies within India's administrative boundary. This global variable is further used in many modules for the further processing, where there is a requirement to obtain the number of grids which has the valid value and within the administrative boundary of India.

Subsequently, the algorithm presented in Pseudo Code 4 is used to determine the index value of valid grids from grid-matrix, which specifically lies within the administrative boundary of India. At first, algorithm identifies the right grid-matrix containing specific number of valid grids (assigned to $Data\_Set\_Number$). As shown in Figure 6.1, index values required for generating the spatial data for any particular grid have been acquired with reference to origin grid. These relative coordinates are calculated in form of row and column indices (i, j) of the grid. The algorithm has been developed to acquire these indices and to save these indices in file that has been named as `INDEX_FILE.csv`. The pseudo code of the algorithm is presented in Pseudo Code 4. File `INDEX_FILE.csv` is used as input for processing the data-set. It also serves as a metadata of grids for IMD precipitation data in India. The metadata stored in the file `INDEX_FILE.csv` was utilized to obtain the separate grid data out of the cluster of the grid-matrix required for the downscaling of the precipitation data. Using the algorithm, the relevant grid data from each year's data file has been extracted. The algorithm is repeatedly applied for the corresponding data of each year data in the temporal duration. IMD precipitation data-set used for validation has temporal duration of 35 years. For every year, total number of grid files on which the algorithm is applied is 1251.

Pseudo Code 3. To determine data set number out of all grids

```
SET Count → 0
SET Data_Set_Number → 0
SET File_List → [IMD Files]
foreach file in File_List do
    foreach DataSet[ROW][COL] in file do
        for row = 1 to ROW do
            for col = 1 to COL do
                if DataSet[row][col] ≥ 0 then
                    Count = Count + 1
                end
            end
        end
        Data_Set_Number = Data_Set_Number + 1
        if Count == Grid_Count then
            GOTO OUTSIDE_LOOP
        end
        Count = 0
    end
end
OUTSIDE_LOOP :
PRINT Data_Set_Number
```

Pseudo Code 4. To determine indices of the grids.

```
CREATE INDEX_FILE
OPEN INDEX_FILE as ifile
SET Count → 0
SET File_List → [IMD Files]
WRITE("X,Y") in ifile
foreach file in File_List do
    Count = 0
    foreach DataSet[ROW][COL] in file do
        Count = Count + 1
        if Count == Data_Set_Number then
            for row = 1 to ROW do
                for col = 1 to COL do
                    if DataSet[row][col] ≥ 0 then
                        WRITE( row, col )in ifile
                    end
                end
            end
            GOTO OUTSIDE_LOOP
        end
    end
end
OUTSIDE_LOOP :
SAVE ifile
CLOSE ifile
```

Pseudo Code 5. To extracting data for the grids.

```
OPEN INDEX_FILE as meta
READ("X,Y") from meta
while READ( x, y ) from meta till EOF do
    STRING Grid_Name = ToString(x,"X",y,"Y.dat")
    CREATE Grid_Name
    OPEN Grid_Name as grid
    OPEN IMD_Year_One_File as year
    foreach DataSet[ROW][COL] in year do
        WRITE( DataSet[x][y] ) in grid
    end
    SAVE grid
    CLOSE grid
    CLOSE year
end
CLOSE meta
```

Using the algorithm discussed in Pseudo Code 5, the files comprising of precipitation data are creating programmatically for 35 years. Each file comprises of precipitation data of a year for a particular grid. Figure 6.1 depicts the directory comprising of these files.

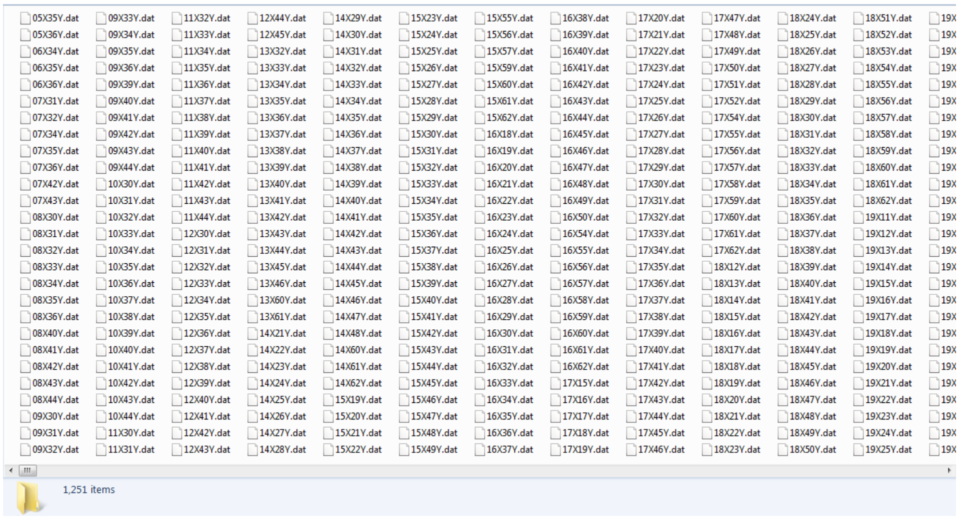


Figure 6.1: Seperated structured grid files from raw data.

Naming of the files has been carried out on basis of grid metadata stored in file `INDEX_FILE.csv`. This compilation of data will results in 43785 files ($= 1251 \times 35$) having precipitation data of the valid grids. For each grid files comprising of precipitation data for each year are saved in 35 number of file directories. For collating the precipitation data of 35 years as a single file for each grid, algorithm as given in Pseudo Code 6 has been developed and implemented. By implementing the algorithm presented in Pseudo Code 6, all grid files in a directory with all year data within it have been retrieved. For the further processing this directory is renamed as `IMD_Grids`.

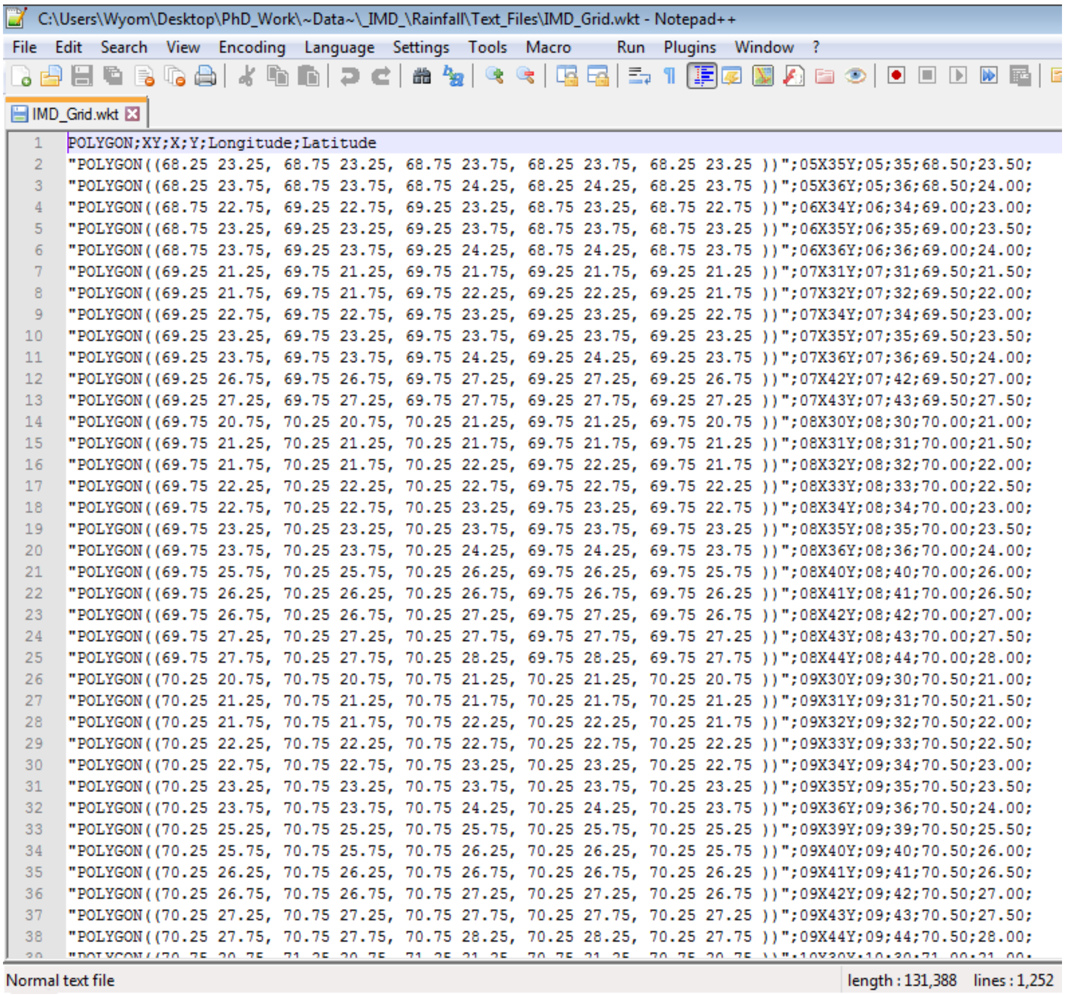
For implementing downscaling, spatial data of the IMD grids as shown in figure 6.2 is also required. In order to generate spatial meta corresponding to each valid grid, base data as given in `INDEX_FILE.csv` is utilized. This spatial metadata is stored in WKT format. WKT format is already discussed in earlier section.

Pseudo Code 6. To collate precipitation data of 35 years for each grid.

```
STRING Parent_Dir
SET Parent_Dir → LOCATEPARENTDIR()
for Count = 1 to Grid_Count do
    OPENDIR Parent_Dir as pdir
    dir_name = READDIR pdir
    OPENDIR dir_name as dir
    for Count = 1 to Grid_Count do
        file_name = READDIR dir
    end
    CLOSEDIR dir
    OPEN file_name as main_file
    while READ( value ) from main_file till EOF do EmptyLoop
    for ( dir_name = READDIR (pdir) ) ≠ NULL do
        OPENDIR dir_name as dir
        for F_Count = 1 to Count do
            file_name = READDIR ( dir )
        end
        OPEN file_name as temp_file
        while READ( value ) from temp_file till EOF do
            WRITE( value ) in main_file
        end
        CLOSE temp_file
        CLOSEDIR dir
    end
    CLOSEDIR pdir
    SAVE main_file
    CLOSE main_file
end
```

For this research, data structure as specified in WKT documentation [ISO/IEC 13249-3: 2016] has been utilized for generating point and polygon vector dataset. However other data structure and other details can be used to extending up the EMSDM. Procedure for creating metadata of spatial component of IMD grids is given in Pseudo Code 7.

The file `IMD_Grid.wkt` which is obtained by aforementioned algorithm is a spatial metadata for IMD grids. Figure 6.2 depicts the basic structure of the `IMD_Grid.wkt` file.




```
POLYGON;XY;X;Y;Longitude;Latitude
"POLYGON((68.25 23.25, 68.75 23.25, 68.75 23.75, 68.25 23.75, 68.25 23.25 ))";05X35Y;05;35;68.50;23.50;
"POLYGON((68.25 23.75, 68.75 23.75, 68.75 24.25, 68.25 24.25, 68.25 23.75 ))";05X36Y;05;36;68.50;24.00;
"POLYGON((68.75 22.75, 69.25 22.75, 69.25 23.25, 68.75 23.25, 68.75 22.75 ))";06X34Y;06;34;69.00;23.00;
"POLYGON((68.75 23.25, 69.25 23.25, 69.25 23.75, 68.75 23.75, 68.75 23.25 ))";06X35Y;06;35;69.00;23.50;
"POLYGON((68.75 23.75, 69.25 23.75, 69.25 24.25, 68.75 24.25, 68.75 23.75 ))";06X36Y;06;36;69.00;24.00;
"POLYGON((69.25 21.25, 69.75 21.25, 69.75 21.75, 69.25 21.75, 69.25 21.25 ))";07X31Y;07;31;69.50;21.50;
"POLYGON((69.25 21.75, 69.75 21.75, 69.75 22.25, 69.25 22.25, 69.25 21.75 ))";07X32Y;07;32;69.50;22.00;
"POLYGON((69.25 22.75, 69.75 22.75, 69.75 23.25, 69.25 23.25, 69.25 22.75 ))";07X34Y;07;34;69.50;23.00;
"POLYGON((69.25 23.25, 69.75 23.25, 69.75 23.75, 69.25 23.75, 69.25 23.25 ))";07X35Y;07;35;69.50;23.50;
"POLYGON((69.25 23.75, 69.75 23.75, 69.75 24.25, 69.25 24.25, 69.25 23.75 ))";07X36Y;07;36;69.50;24.00;
"POLYGON((69.25 26.75, 69.75 26.75, 69.75 27.25, 69.25 27.25, 69.25 26.75 ))";07X42Y;07;42;69.50;27.00;
"POLYGON((69.25 27.25, 69.75 27.25, 69.75 27.75, 69.25 27.75, 69.25 27.25 ))";07X43Y;07;43;69.50;27.50;
"POLYGON((69.75 20.75, 70.25 20.75, 70.25 21.25, 69.75 21.25, 69.75 20.75 ))";08X30Y;08;30;70.00;21.00;
"POLYGON((69.75 21.25, 70.25 21.25, 70.25 21.75, 69.75 21.75, 69.75 21.25 ))";08X31Y;08;31;70.00;21.50;
"POLYGON((69.75 21.75, 70.25 21.75, 70.25 22.25, 69.75 22.25, 69.75 21.75 ))";08X32Y;08;32;70.00;22.00;
"POLYGON((69.75 22.25, 70.25 22.25, 70.25 22.75, 69.75 22.75, 69.75 22.25 ))";08X33Y;08;33;70.00;22.50;
"POLYGON((69.75 22.75, 70.25 22.75, 70.25 23.25, 69.75 23.25, 69.75 22.75 ))";08X34Y;08;34;70.00;23.00;
"POLYGON((69.75 23.25, 70.25 23.25, 70.25 23.75, 69.75 23.75, 69.75 23.25 ))";08X35Y;08;35;70.00;23.50;
"POLYGON((69.75 23.75, 70.25 23.75, 70.25 24.25, 69.75 24.25, 69.75 23.75 ))";08X36Y;08;36;70.00;24.00;
"POLYGON((69.75 25.75, 70.25 25.75, 70.25 26.25, 69.75 26.25, 69.75 25.75 ))";08X40Y;08;40;70.00;26.00;
"POLYGON((69.75 26.25, 70.25 26.25, 70.25 26.75, 69.75 26.75, 69.75 26.25 ))";08X41Y;08;41;70.00;26.50;
"POLYGON((69.75 26.75, 70.25 26.75, 70.25 27.25, 69.75 27.25, 69.75 26.75 ))";08X42Y;08;42;70.00;27.00;
"POLYGON((69.75 27.25, 70.25 27.25, 70.25 27.75, 69.75 27.75, 69.75 27.25 ))";08X43Y;08;43;70.00;27.50;
"POLYGON((69.75 27.75, 70.25 27.75, 70.25 28.25, 69.75 28.25, 69.75 27.75 ))";08X44Y;08;44;70.00;28.00;
"POLYGON((70.25 20.75, 70.75 20.75, 70.75 21.25, 70.25 21.25, 70.25 20.75 ))";09X30Y;09;30;70.50;21.00;
"POLYGON((70.25 21.25, 70.75 21.25, 70.75 21.75, 70.25 21.75, 70.25 21.25 ))";09X31Y;09;31;70.50;21.50;
"POLYGON((70.25 21.75, 70.75 21.75, 70.75 22.25, 70.25 22.25, 70.25 21.75 ))";09X32Y;09;32;70.50;22.00;
"POLYGON((70.25 22.25, 70.75 22.25, 70.75 22.75, 70.25 22.75, 70.25 22.25 ))";09X33Y;09;33;70.50;22.50;
"POLYGON((70.25 22.75, 70.75 22.75, 70.75 23.25, 70.25 23.25, 70.25 22.75 ))";09X34Y;09;34;70.50;23.00;
"POLYGON((70.25 23.25, 70.75 23.25, 70.75 23.75, 70.25 23.75, 70.25 23.25 ))";09X35Y;09;35;70.50;23.50;
"POLYGON((70.25 23.75, 70.75 23.75, 70.75 24.25, 70.25 24.25, 70.25 23.75 ))";09X36Y;09;36;70.50;24.00;
"POLYGON((70.25 25.25, 70.75 25.25, 70.75 25.75, 70.25 25.75, 70.25 25.25 ))";09X39Y;09;39;70.50;25.50;
"POLYGON((70.25 25.75, 70.75 25.75, 70.75 26.25, 70.25 26.25, 70.25 25.75 ))";09X40Y;09;40;70.50;26.00;
"POLYGON((70.25 26.25, 70.75 26.25, 70.75 26.75, 70.25 26.75, 70.25 26.25 ))";09X41Y;09;41;70.50;26.50;
"POLYGON((70.25 26.75, 70.75 26.75, 70.75 27.25, 70.25 27.25, 70.25 26.75 ))";09X42Y;09;42;70.50;27.00;
"POLYGON((70.25 27.25, 70.75 27.25, 70.75 27.75, 70.25 27.75, 70.25 27.25 ))";09X43Y;09;43;70.50;27.50;
"POLYGON((70.25 27.75, 70.75 27.75, 70.75 28.25, 70.25 28.25, 70.25 27.75 ))";09X44Y;09;44;70.50;28.00;
```


Figure 6.2: Seperated structured grid files from raw data.

Pseudo Code 7. To creating spatial metadata for IMD grids.

```
SET GRID_START_LONG → START_LONG
SET GRID_START_LAT → START_LAT
SET OFFSET → 0.5
SET C_OFFSET → 0.25
CREATE IMD_Grid.wkt
OPEN IMD_Grid.wkt as s_meta
OPEN INDEX_FILE as meta
READ("X,Y") from meta
WRITE("POLYGON;XY;X;Y;Longitude;Latitude") in s_meta
while READ( x, y ) from meta till EOF do
    Min_Long = GRID_START_LONG + OFFSET ∗ ( x - 1 )
    Max_Long = Min_Long + OFFSET
    Min_Lat = GRID_START_LAT + OFFSET ∗ ( y - 1 )
    Max_Lat = Min_Lat + OFFSET
    C_Long = Min_Long + C_OFFSET
    C_Lat = Min_Lat + C_OFFSET
    WRITE( "\n"POLYGON((" ) in s_meta
    WRITE( Min_Long, Min_Lat, ",” ) in s_meta
    WRITE( Max_Long, Min_Lat, "," ) in s_meta
    WRITE( Max_Long, Max_Lat, "," ) in s_meta
    WRITE( Min_Long, Max_Lat, "," ) in s_meta
    WRITE( Min_Long, Min_Lat, "))\";" ) in s_meta
    WRITE( x"X"y"Y;" ) in s_meta
    WRITE( x, ";", y, ";" ) in s_meta
    WRITE( C_Long, ";", C_Lat, ";" ) in s_meta
end
SAVE s_meta
CLOSE s_meta
CLOSE meta
```

The metadata file viz. `IMD_Grid.wkt` was used for generating shapefiles data. Quantum GIS (QGIS) was used for generation of spatial layer using WKT format Following procedure has been developed and implemented for importing WKT file as spatial layer in QGIS. WKT file is referenced to WGS84 projection system:

Open Qgis

Goto Layer

Goto Add Layer

Goto Add Delimite Text Layer

Browse WKT file

Set File Format as Custom Delimiters

Set delimiters as

Semicolon

Quote "

Escape "

Record Options

Number of header lines to discard 0

First record has field names checked

Geometry definition

Well known text (WKT) checked

Geometry field POLYGON

Press OK

Coordinate reference System Selector

Coordinate reference System WGS84

Press OK

Moreover, WKT file can also be converted to shapefile format using export tool of QGIS software. Shapefile is one of the most commonly used spatial data format. Following procedure was developed and implemented for converting the spatial layer generated from aforementioned layer to shapefile format:

In Layer Panel

Right Click on Layer

Save As

In Window Save vector layer as...

Format *ESRI Shapefile*

Save as *set File_Name*

CRS *EPSG:4326, WGS 84*

Press OK

Moreover if the user does not have software which can support the WKT file format, then user can use shape-file for processing and subsequently analysis. For example, GIS software namely DIVA GIS software does not support WKT file format. Hence user of DIVA GIS can use shape-file format for visual analysis.

6.4 Pre-Processing of GCM Data-sets (Predictors)

6.4.1 GCM Data-Set (CanESM2)

Besides, precipitation data-set from IMD, GCM data-set has been required for preprocessing. As discussed in earlier section, CanESM2 as GCM has been used for validating the EMSDM. CanESM2 stands for second generation of Canadian Earth System Model. Canadian Centre for Climate Modelling and Analysis (CCCma) of Environment Canada developed this climate model. CanESM2 is the 4th generation coupled global climate model. In IPCC 5^{th} Assessment Report (AR5) this climate model is contributed by CCCma.

6.4.2 Preprocessing of CanESM2

CanESM2 is divided in 128×64 grid cells to cover global domain. Spatial resolution of each grid is $\sim 2.7906^o \times 2.8125^o$, which is uniform in along longitude and along latitude it is with a rough resolution of 2.7906^o. Each grid's data is stored in a folder named BOX_iiiX_jjY where iii and jj represent the longitudinal and latitudinal indices. The index detail or grid definition with the longitude and latitude of the centre of grid were obtained from Canadian Centre for Climate Modelling and Analysis (CCCma) website [CCCma, 2018].

6.4.2.1 Generation of Spatial Metadata for GCM Grids

`GCM_CanESM2_X_Longitude.csv` and `GCM_CanESM2_Y_Latitude.csv` as in tables 6.3 and 6.4 are used to provide the grid index information of GCM box. These two files are used to create a spatial metadata file `GCM_CanESM2_Ploygon.wkt`. Using this algorithm, user can generate this spatial metadata file with the help of Longitude Index metadata file and Latitude Index metadata file. File obtained by this algorithm comprises of the grids covering the entire Earth. The algorithm to create the spatial metadata file is given in Pseudo Code 8.

Table 6.3: Index of X and its Corresponding Longitude

Index of X and its Corresponding Longitude							
X(iii)	**Longitude**	**X(iii)**	**Longitude**	**X(iii)**	**Longitude**	**X(iii)**	**Longitude**
001	000.0000	**033**	090.0000	**065**	180.0000	**097**	270.0000
002	002.8125	**034**	092.8125	**066**	182.8125	**098**	272.8125
003	005.6250	**035**	095.6250	**067**	185.6250	**099**	275.6250
004	008.4375	**036**	098.4375	**068**	188.4375	**100**	278.4375
005	011.2500	**037**	101.2500	**069**	191.2500	**101**	281.2500
006	014.0625	**038**	104.0625	**070**	194.0625	**102**	284.0625
007	016.8750	**039**	106.8750	**071**	196.8750	**103**	286.8750
008	019.6875	**040**	109.6875	**072**	199.6875	**104**	289.6875
009	022.5000	**041**	112.5000	**073**	202.5000	**105**	292.5000
010	025.3125	**042**	115.3125	**074**	205.3125	**106**	295.3125
011	028.1250	**043**	118.1250	**075**	208.1250	**107**	298.1250
012	030.9375	**044**	120.9375	**076**	210.9375	**108**	300.9375
013	033.7500	**045**	123.7500	**077**	213.7500	**109**	303.7500
014	036.5625	**046**	126.5625	**078**	216.5625	**110**	306.5625
015	039.3750	**047**	129.3750	**079**	219.3750	**111**	309.3750
016	042.1875	**048**	132.1875	**080**	222.1875	**112**	312.1875
017	045.0000	**049**	135.0000	**081**	225.0000	**113**	315.0000
018	047.8125	**050**	137.8125	**082**	227.8125	**114**	317.8125
019	050.6250	**051**	140.6250	**083**	230.6250	**115**	320.6250
020	053.4375	**052**	143.4375	**084**	233.4375	**116**	323.4375
021	056.2500	**053**	146.2500	**085**	236.2500	**117**	326.2500
022	059.0625	**054**	149.0625	**086**	239.0625	**118**	329.0625
023	061.8750	**055**	151.8750	**087**	241.8750	**119**	331.8750
024	064.6875	**056**	154.6875	**088**	244.6875	**120**	334.6875
025	067.5000	**057**	157.5000	**089**	247.5000	**121**	337.5000
026	070.3125	**058**	160.3125	**090**	250.3125	**122**	340.3125
027	073.1250	**059**	163.1250	**091**	253.1250	**123**	343.1250
028	075.9375	**060**	165.9375	**092**	255.9375	**124**	345.9375
029	078.7500	**061**	168.7500	**093**	258.7500	**125**	348.7500
030	081.5625	**062**	171.5625	**094**	261.5625	**126**	351.5625
031	084.3750	**063**	174.3750	**095**	264.3750	**127**	354.3750
032	087.1875	**064**	177.1875	**096**	267.1875	**128**	357.1875

Table 6.4: Index of Y and its corresponding Latitude

Index of Y and its corresponding Latitude							
Y(jj)	**Latitude**	**Y(jj)**	**Latitude**	**Y(jj)**	**Latitude**	**Y(jj)**	**Latitude**
01	-87.863	**17**	-43.254	**33**	01.395	**49**	46.044
02	-85.096	**18**	-40.463	**34**	04.185	**50**	48.835
03	-82.312	**19**	-37.673	**35**	06.976	**51**	51.625
04	-79.525	**20**	-34.882	**36**	09.767	**52**	54.416
05	-76.736	**21**	-32.091	**37**	12.557	**53**	57.206
06	-73.947	**22**	-29.301	**38**	15.348	**54**	59.997
07	-71.157	**23**	-26.51	**39**	18.138	**55**	62.787
08	-68.367	**24**	-23.72	**40**	20.929	**56**	65.577
09	-65.577	**25**	-20.929	**41**	23.72	**57**	68.367
10	-62.787	**26**	-18.138	**42**	26.51	**58**	71.157
11	-59.997	**27**	-15.348	**43**	29.301	**59**	73.947
12	-57.206	**28**	-12.557	**44**	32.091	**60**	76.736
13	-54.416	**29**	-09.767	**45**	34.882	**61**	79.525
14	-51.625	**30**	-06.976	**46**	37.673	**62**	82.312
15	-48.835	**31**	-04.185	**47**	40.463	**63**	85.096
16	-46.044	**32**	-01.395	**48**	43.254	**64**	87.863

Pseudo Code 8. To creating spatial metadata for GCM grids.

```
OPEN GCM_CanESM2_X_Longitude.txt as meta_x
OPEN GCM_CanESM2_Y_Latitude.txt as meta_y
CREATE GCM_CanESM2_Polygon.wkt
OPEN GCM_CanESM2_Polygon.wkt as s_meta
WRITE("POLYGON;XY;X;Y;Longitude;Latitude") in s_meta
READ( X, Min_Long ) from meta_x
while READ( X, Max_Long ) from meta_x till EOF do
    C_Long = Min_Long + 0.5 * ( Max_Long - Min_Long )
    REWIND meta_y
    READ( Y, Min_Lat ) from meta_y
    while READ( Y, Max_Lat ) from meta_y till EOF do
        C_Lat = Min_Lat + 0.5 * ( Max_Lat - Min_Lat )
        WRITE( "\n"POLYGON((" ) in s_meta
        WRITE( Min_Long, Min_Lat, ",” ) in s_meta
        WRITE( Max_Long, Min_Lat, "," ) in s_meta
        WRITE( Max_Long, Max_Lat, "," ) in s_meta
        WRITE( Min_Long, Max_Lat, "," ) in s_meta
        WRITE( Min_Long, Min_Lat, "))\";" ) in s_meta
        WRITE( x"X"y"Y;" ) in s_meta
        WRITE( x, ";", y, ";" ) in s_meta
        WRITE( C_Long, ";", C_Lat, ";" ) in s_meta
        Min_Lat = Max_Lat
    end
    Min_Long = Max_Long
end
SAVE s_meta
CLOSE s_meta
CLOSE meta_y
CLOSE meta_x
```

6.4.2.2 Generation of Metadata for GCM Grids of AOI

Algorithm from Pseudo Code 8 produces grids for the world. For EMSDM, CanESM2 grids that cover India are required, which are obtained by spatial overlay analysis of `IMD_Grid.wkt` and `GCM_CanESM2_Polygon.wkt` data-sets. This spatial overlay analysis is carried out under QGIS environment. The procedure is automated in the EMSDM and used for extracting the relevant CanESM2 grids is discussed below:

Open Qgis

Goto Layer

Goto Add Layer

Goto Add Delimite Text Layer

Browse GCM_ CanESM2_ Polygon.wkt file

Press OK

Coordinate reference System Selector

Coordinate reference System WGS84

Press OK Goto Layer

Goto Add Layer

Goto Add Delimite Text Layer

Browse GCM_ CanESM2_ Polygon.wkt file

Press OK

Coordinate reference System Selector

Coordinate reference System WGS84

Press OK Goto Vector

Goto Research Tools

Goto Select by location

Set Layer to Select From: GCM_ CanESM2_ Polygon

Set Additional layer : IMD_ Grid

Geometric Predicate : *Overlaps*

Modify current Selection by : *Creating new Selection*

Press RUN

In Layer Panel

Select *GCM_ CanESM2_ Polygon*

Right *Click on Layer*

Save *As*

In Window Save vector layer as...

Format : *Comma Separated Value [CSV]*

Save as : *set File_ Name*

CRS : *EPSG:4326, WGS 84*

Encoding: *Save Only Selected Features*

Press OK

Using the procedure discussed above, the relevant CanESM2 grids that corresponds to India have been obtained. These grids data was stored as `GCM_CanESM2_Polygon_India.csv`. This file comprises of the information of the grids that are required for downscaling of India.

6.4.2.3 Extraction of Spatial Metadata for GCM Grids of AOI

The metadata created in section 6.4.2.2 is used to filter out the required grid from GCM_CanESM2_Polygon.wkt for generation of spatial metadata of CanESM2 grids. The algorithm presented in Pseudo Code 9 is used to segregate the spatial metadata of AOI from the spatial metadata generated through the algorithms and procedures discussed in section 6.4.2.1. This segregation process is carried out using the spatial metadata for GCM grids and metadata for GCM grids of AOI. These metadata files are obtained from the algorithms and procedures discussed in section 6.4.2.1 and section 6.4.2.2.

Pseudo Code 9. To extract spatial metadata of GCM grids for AOI.

```
SET Run_It → TRUE
OPEN GCM_Name_Polygon.wkt as s_meta
OPEN GCM_Name_Polygon_aoi.csv as meta
CREATE GCM_Name_Polygon_aoi.wkt
OPEN GCM_Name_Polygon_aoi.wkt as s_meta_aoi
READ("POLYGON;XY;X;Y;Longitude;Latitude") from s_meta
READ("XY") from meta
WRITE("POLYGON;XY;X;Y;Longitude;Latitude") in s_meta_aoi
while READ( x"X"y"Y" ) from meta till EOF do
    while Run_It do
        READ( "\n"POLYGON((" ) from s_meta
        READ( Min_Long, Min_Lat, "," ) from s_meta
        READ( Max_Long, Min_Lat, "," ) from s_meta
        READ( Max_Long, Max_Lat, "," ) from s_meta
        READ( Min_Long, Max_Lat, "," ) from s_meta
        READ( Min_Long, Min_Lat, "))\";" ) from s_meta
        READ( x"X"y"Y;" ) from s_meta
        READ( x, ";", y, ";" ) from s_meta
        Run_It = READ( C_Long, ";", C_Lat, ";" ) from s_meta
        if x == X && y == Y then
            WRITE( "\n"POLYGON((" ) in s_meta_aoi
            WRITE( Min_Long, Min_Lat, "," ) in s_meta_aoi
            WRITE( Max_Long, Min_Lat, "," ) in s_meta_aoi
            WRITE( Max_Long, Max_Lat, "," ) in s_meta_aoi
            WRITE( Min_Long, Max_Lat, "," ) in s_meta_aoi
            WRITE( Min_Long, Min_Lat, "))\";" ) in s_meta_aoi
            WRITE( x"X"y"Y;" ) in s_meta_aoi
            WRITE( x, ";", y, ";" ) in s_meta_aoi
            WRITE( C_Long, ";", C_Lat, ";" ) in s_meta_aoi
            BREAK
        end
    end
end
SAVE s_meta_aoi
CLOSE s_meta_aoi
CLOSE s_meta
CLOSE meta
```

6.4.2.4 Extraction of GCM Grid Data-Sets for AOI

When downscaling is performed for large spatial extent (AOI), various GCM grids are associated with AOI. For obtaining the GCM grids data-set from the official data site, there is a need of automation of this process. Pseudo Code 10 is used to create a URL file. This file is used by the EMSDM to automatically download the required GCM parametric data in form of zip files from the official data site.

Pseudo Code 10. To create url file to obtaine GCM Grids data-sets.

```
SET Part_URL
SET Full_URL
OPEN GCM_CanESM2_Polygon_aoi.csv as meta
CREATE GCM_CanESM2_Polygon_aoi.url
OPEN GCM_CanESM2_Polygon_aoi.url as data_urls
READ("XY") from meta
while READ( x"X"y"Y" ) from meta till EOF do
    Full_URL = ToString(Part_URL, x, "X_", y, "Y.zip" )
    WRITE( Full_URL ) in data_urls
end
SAVE data_urls
CLOSE data_urls
CLOSE meta
```

Pseudo Code 11. To find GCM grid corresponding to a predictand grid.

```
Set Run_It → TRUE
Set DATA_SOURCE → "IMD"
Set AOI_NAME → "India"
Set GCM_NAME → "CanESM2"
Set GRID_NAME → "001X001Y"
Set GRID_START_LONG → 66.75
Set GRID_START_LAT → 6.75
Set OFFSET → 0.5
Set C_OFFSET → 0.25
String WKT_File
Write( "..\\Text_Files\\GCM_" GCM_NAME "_" AOI_Name ".csv" ) in WKT_File
Open WKT_File as meta
Read( Grid_X "X" Grid_Y "Y" ) from GRID_NAME
Grid_Long = GRID_START_LONG + 0.5 ∗ ( Grid_X - 1 )
Grid_Lat = GRID_START_LAT + 0.5 ∗ ( Grid_Y - 1 )
Read("POLYGON;XY;X;Y;Longitude;Latitude") from meta
while Run_It do
    Read( "\n"POLYGON((" ) from meta
    Read( Min_Long, Min_Lat, "," ) from meta
    Read( Max_Long, Min_Lat, "," ) from meta
    Read( Max_Long, Max_Lat, "," ) from meta
    Read( Min_Long, Max_Lat, "," ) from meta
    Read( Min_Long, Min_Lat, "))\";" ) from meta
    Read( x"X"y"Y;" ) from meta
    Read( x, ";", y, ";" ) from meta
    Run_It = Read( C_Long, ";", C_Lat, ";" ) from meta
    if Grid_Long ≥ Min_Long then
        if Grid_Long < Max_Long then
            if Grid_Lat ≥ Min_Lat then
                if Grid_Lat < Max_Lat then
                    Write( "BOX_" x "X_" y "Y" ) in GCM_Grid_Name
                    Break
                end
            end
        end
    end
end
Close meta
Print GCM_Grid_Name
```

Pseudo Code 12. To Generate Monthly GCM Grid Data-Set.

```
STRING Temporal_Type, GCM_Dir, Box_Name, File_Name
STRING File_Read_Dir, File_Write_Dir
STRING File_Read_Path, _Write_Path
SET Source_Dir → "daily"
SET Target_Dir[] → { "monthly", "annual", "decadal" }
SET GCM_Dir → LOCATE-GCM-DIR()
SET GCM_SubDir[ ] → GET-GCM-SUBDIR()
Temporal_Type = GET-ENTRY( "Enter Required Temporal Type" )
if Temporal_ Type == "monthly" then n_tt = 1
if Temporal_ Type == "annual" then n_tt = 2
if Temporal_ Type == "decadal" then n_tt = 3
OPENDIR GCM_Dir as gcm_dir
while ( Box_ Name = READDIR ( gcm_ dir ) ) != NULL do
    for g_ i = 0 to n_ GCM_ SubDir do
        WRITE( "..\\data\\GCM\\" Source_Dir "\\" GCM_Name "\\" Box_Name "\\"
          GCM_SubDir[g_i]) in File_Read_Dir
        WRITE( "..\\data\\GCM\\" Target_Dir[n\_tt] "\\" GCM_Name "\\" Box_Name
          "\\" GCM_SubDir[g_i]) in File_Write_Dir
        OPENDIR File_Read_Dir as file_dir
        while ( File_ Name = READDIR ( Box_Name ) ) != NULL do
            WRITE( File_Read_Dir "\\" File_Name ) in File_Read_Path
            WRITE( File_Write_Dir "\\" File_Name ) in File_Write_Path
            OPEN File_Read_Path as read_file
            CREATE File_Write_Path
            OPEN File_Write_Path as write_file
            if Temporal_ Type == "monthly" then
                CONVERTTOMONTH( read_file, write_file )
            end
            if Temporal_ Type == "annual" then
                CONVERTTOANNUAL( read_file, write_file )
            end
            if Temporal_ Type == "decadal" then
                CONVERTTODECADAL( read_file, write_file )
            end
            SAVE write_file
            CLOSE write_file
            CLOSE read_file
        end
        CLOSEDIR file_dir
    end
end
CLOSEDIR gcm_dir
```

Pseudo Code 13. To evaluate Pearson's correlation coefficient using.

```
String Correlation_Type
String Data_Source
String Climate_Variable
String AOI_Name
String GCM_Name
String Temporal_Type
String Input_File
String Grid_Name
String GCM_Grid_Dir
String Station_Data_Path
String Correlation_Results_Data_Path
String gcm_grid_found → FALSE
Data_Source = Get-Entry( "Enter Name of Data Source" )
AOI_Name = Get-Entry( "Enter Name of AOI" )
GCM_Name = Get-Entry( "Enter Name of GCM" )
Climate_Variable = Get-Entry( "Enter Name of Climate Variable" )
Correlation_Type = Get-Entry( "Enter Type of Correlation Type Required" )
if Correlation_Type == "parametric" then
    n_ctype = 1
end
if Correlation_Type == "non-parametric" then
    n_ctype = 2
end
Temporal_Type = Get-Entry( "Enter Required Temporal Type" )
Write( "..\\Text_Files\\" Data_Source "_" AOI_Name "_Index.csv" ) in Input_File
Open Input_File as meta_xy
Read( "XY" ) from meta_xy
while Read( Grid_Name ) from meta_xy till EOF do
    gcm_grid_found = Find_GCM_Grid( Data_Source, AOI_Name, GCM_Name, Grid_Name )
    if gcm_grid_found then
        if n_ctype == 1 then
            Write( "..\\data\\" Data_Source "\\" Climate_Variable "\\" AOI_Name "\\"
              Temporal_Type "\\" Grid_Name ".csv" ) in Station_Data_Path
            Write(  "..\\Output\\" GCM_Name "\\" Data_Source "\\" Climate_Variable "\\"
              AOI_Name "\\" Temporal_Type "\\correlation\\parametric\\" Grid_Name ".csv"  ) in
              Correlation_Results_Data_Path
            Write( "..\\data\\GCM\\" Temporal_Type "\\" GCM_Name "\\" _GCM_GRID_ ) in
              GCM_Grid_Dir
            Correlation_Pearson( Station_Data_Path, Correlation_Results_Data_Path,
              GCM_Grid_Dir )
        end
        if n_ctype == 2 then
            Write( "..\\data\\nonparametric\\" Data_Source "\\" Climate_Variable "\\" AOI_Name
              "\\" Temporal_Type "\\" Grid_Name ".csv" ) in Station_Data_Path
            Write(  "..\\Output\\" GCM_Name "\\" Data_Source "\\" Climate_Variable "\\"
              AOI_Name "\\" Temporal_Type "\\correlation\\nonparametric\\" Grid_Name ".csv"  ) in
              Correlation_Results_Data_Path
            CreateRankIndex(  GCM_Name, _GCM_GRID_ )
            Write( "..\\data" Corr\_Type "\\GCM\\" Temporal_Type "\\" GCM_Name "\\"
              _GCM_GRID_ ) in GCM_Grid_Dir
            Correlation_Spearman( Station_Data_Path, Correlation_Results_Data_Path,
              GCM_Grid_Dir )
        end
    end
end
Close meta_xy
```

Pseudo Code 14. To Downscaled Model using MLR and MvLR.

```
String Regression_Type
String Data_Source
String Temporal_Type
String Climate_Variable
String AOI_Name, GCM_Name, Grid_Name
String Index_File, Input_File, Output_File
Set gcm_grid_found → FALSE
Set Sum → 0
Set Count → 0
Data_Source = Get-Entry( "Enter Name of Data Source" )
AOI_Name = Get-Entry( "Enter Name of AOI" )
GCM_Name = Get-Entry( "Enter Name of GCM" )
Climate_Variable = Get-Entry( "Enter Name of Climate Variable" )
Regression_Type = Get-Entry( "Enter Type of Regression Required" )
if Correlation_Type == "MLR" then
    n_rtype = 1
end
if Correlation_Type == "MvLR" then
    n_rtype = 2
end
Temporal_Type = Get-Entry( "Enter Required Temporal Type" )
Write( "..\\Text_Files\\" Data_Source "_" AOI_Name "_Index.csv" ) in Index_File
Open Index_File as meta_xy
Read( "XY" ) from meta_xy
while Read( Grid_Name ) from meta_xy till EOF do
    gcm_grid_found = Find_GCM_Grid( Data_Source, AOI_Name, GCM_Name, Grid_Name )
    if gcm_grid_found then
        if n_rtype == 1 then
            Write( "..\\Output\\" GCM\_Name "\\" Data_Source "\\" Climate_Variable "\\"
              AOI_Name "\\" Temporal_Type "\\correlation\\parametric\\" Grid_Name ".csv" ) in
              Input_File
            Write( "..\\Output\\" GCM_Name "\\" Data_Source "\\" Climate_Variable "\\"
              AOI_Name "\\" Temporal_Type "\\regression\\MLR\\" Grid_Name ".csv" ) in
              Output_File
            if Regression-MLR ( Input_File, Output_File ) then
                Count = Count + 1
            end
            Sum = Sum + 1
        end
        if n_rtype == 2 then
            n_Group = Get-Entry( "Enter Number of Regression Group" )
            Write( "..\\Output\\" GCM\_Name "\\" Data_Source "\\" Climate_Variable "\\"
              AOI_Name "\\" Temporal_Type "\\correlation\\parametric\\" Grid_Name ".csv" ) in
              Input_File
            Write( "..\\Output\\" GCM_Name "\\" Data_Source "\\" Climate_Variable "\\"
              AOI_Name "\\" Temporal_Type "\\regression\\MvLR\\" Grid_Name ".csv" ) in
              Output_File
            if Regression-MvLR ( Input_File, Output_File, n_Group ) then
                Count = Count + 1
            end
            Sum = Sum + 1
        end
    end
end
Print (Count/Sum) × 100 % Success Rate
Close meta_xy
```

6.5 Spatial Analysis

Spatial Analysis has been carried out to spatially link the local grid and the GCM grid. The algorithm required for carrying out spatial analysis is presented in Pseudo Code 11. The inputs required for spatial analysis are local grid data source, AOI name, GCM name, and grid id. The algorithm uses the aforementioned inputs to carry out the spatial overlay analysis in order to map the GCM grid with the local station grid.

6.6 Temporal Analysis

Temporal Analysis has been carried out to temporally map the local grid data-set and the it's corresponding GCM grid data-set which is obtained from spatial analysis performed in previous section. The algorithm required to carrying out temporal analysis is presented in Pseudo Code 12.

Temporal mapping of local grid data-set and GCM grid data-set has been carried out. Moreover, in to reduce the time complexity of the computational process for the further processing, temporal transformation of the local grid data-set as well as GCM grid data-set is also been carried out using Pseudo Code 12.

6.7 Screening Predictors

As discussed in the section 5.2.4, in this step, selection of set of dominating predictors for downscaling process has been carried out using the algorithm presented in Pseudo Code 13.

In Pseudo Code 13, screening of predictors is carried out by using parametric and non-parametric statistics. As discussed in chapter 4, while Spearman's Correlation Statistics has been used as non-parametric statistics, Pearson Correlation Statistics has been used as parametric statistics. Pseudo Code 13 in-

ternally uses the GCM output dataset of the reference grids of specified AOI to compute statistics and applies the criteria discussed in section 5.2.4 to screen out the GCM parameters having the correlation less the user defined threshold. Moreover user can provide the weightage file for the predictors for carry out downscaling using the specified set of predictors, to analyse the affect of these predictors on climate change.

6.8 Model Generation

Mathematical model solution discussed in the section 5.2.4, is implemented as algorithm for development of statistical downscaling model. This algorithm is presented in Pseudo Code 14. The algorithm automates the model development for the all the grids of the AOI. The inputs required for the development of the model are specified in lines 1-6 of the pseudo code.

6.9 Concluding Remarks

In this chapter algorithms and procedures required for implementing the EMSDM has been discussed in detail. Algorithms and procedures of EMSDM are programmed in C language. Henceforth EMSDM can be implemented irrespective of the Operating Systems viz. Windows, Unix, Linux, macOS etc and possess advantage of interoperability over widely used Window Based SDSM software. Moreover, in addition of carrying out statistical downscaling, EMSDM generates the interoperable spatial data for corresponding GCM and local grids. These spatial data can be used for different climate studies. After discussion on implementation of EMSDM in this chapter, in Chapter 7, applicability of EMSDM is demonstrated using CanESM2 as GCM data-set, IMD as local data-set and India as a selected AOI.

Chapter 7

Application of EMSDM

7.1 General

In this chapter, the application of EMSDM is is presented to demonstrate its applicability for the statistical downscaling for the given area of interest.

Local precipitation dataset of India acquired from Indian Meteorological Department (IMD) of 1241 spatial grids and corresponding GCM dataset comprising of 60 spatial grids for generation of results. IMD data set is avaliable as a file with naming format as **rf**xx_yyyy**.grd**, where xx denotes the resolution of grid in degrees and $yyyy$ denotes the year of the mesurements of the precipitation observations. As a primary climate data, precipitation data acquired from IMD is structured as the grids having resolution of $0.5° \times 0.5°$. Temporal resolution of IMD data-set is one day. Temporal duration of the precipitation data-set is from year 1971 to 2005.

As a GCM, CanESM2 is acquired from CCCma is also structured as the grids with grid resolution $2.7906° \times 2.8125°$. Temporal resolution of predictors data-set obtained from this GCM is taken as one day. Temporal duration of the CanESM2 data-set is from year 1961 to 2005 for historical and NCEP/NCAR reanalysis data and from year 2006 to 2100 for scenarios RCP26, RCP45 and RCP85.

7.2 Pre-Processing of Local Data-sets (Predictand)

Local data are preprocessed using the algorithms presented in sections 6.3 of Chapter 6. Spatial indices of the local stations are encoded in proper format. As shown in Figure 7.1, file comprising of spatial indices of the local stations has been generated using the algorithm presented in section 6.3. File comprising of these spatial indices in CSV file format are generated programmatically and stored in programmatically generated directory which are depicted in Figure 7.2.

As shown in Figure 7.3, spatial metadata for the IMD grids in WKT format are generated using the algorithms presented in section 6.3 and the indices depicted in Figure 7.1.

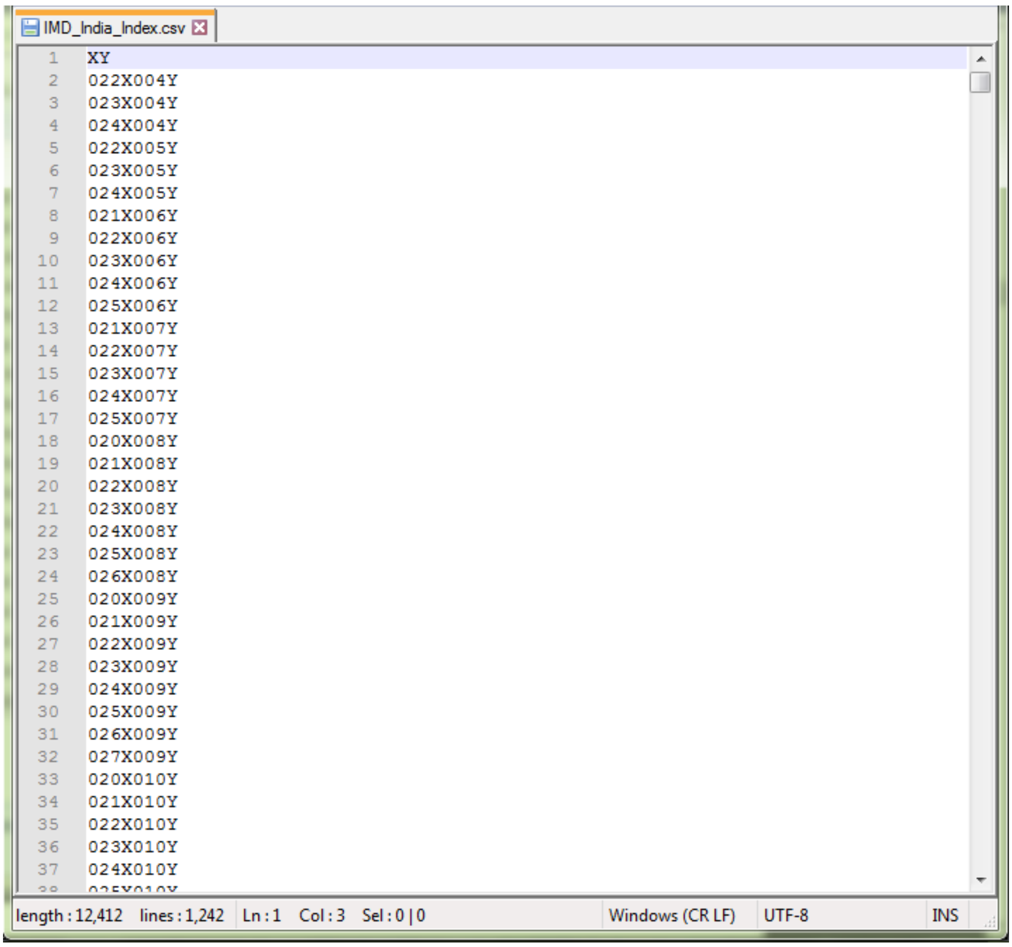


Figure 7.1: IMD-Index

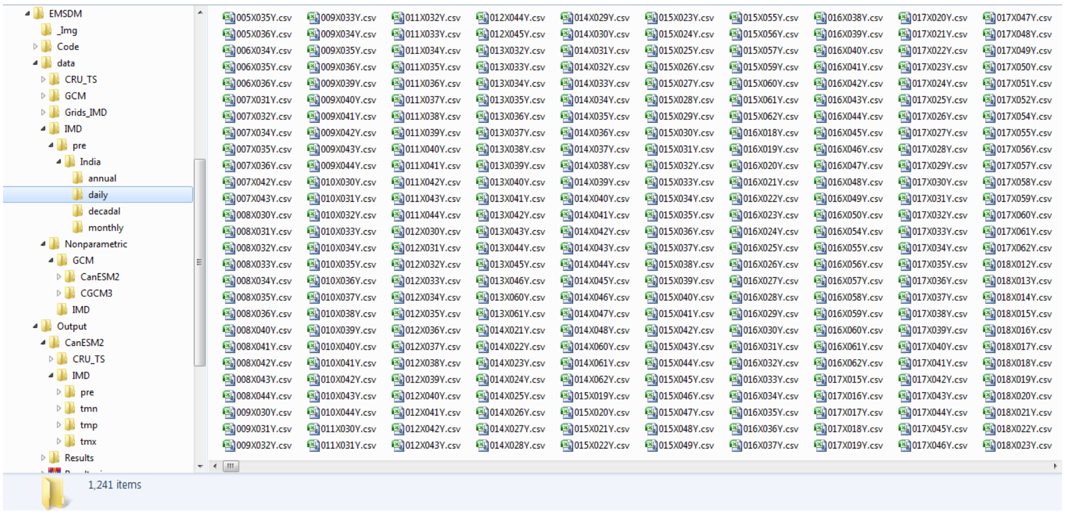


Figure 7.2: IMD-Grids

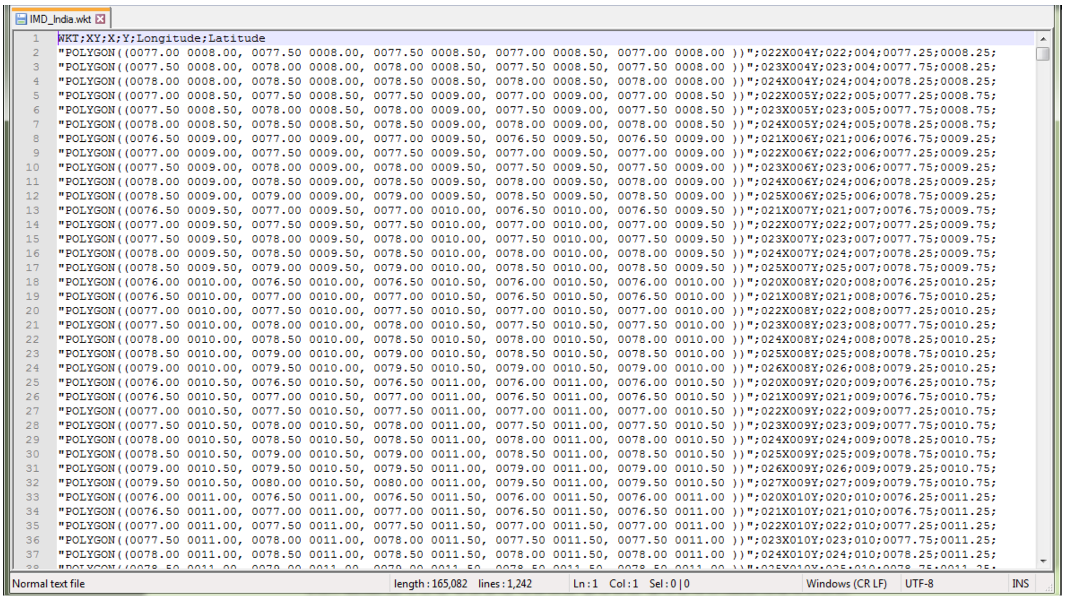


Figure 7.3: IMD Spatial Metadata

The geo-visualization of the spatial metadata for the IMD grids in QGIS software is depicted in Figure 7.4, Figure 7.5 and Figure 7.6.

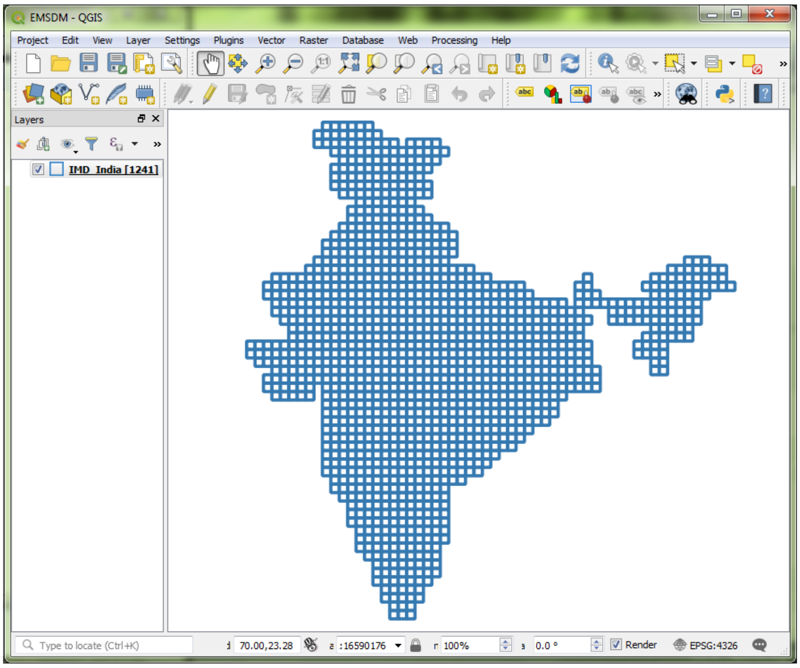


Figure 7.4: IMD Spatial Metadata in QGIS

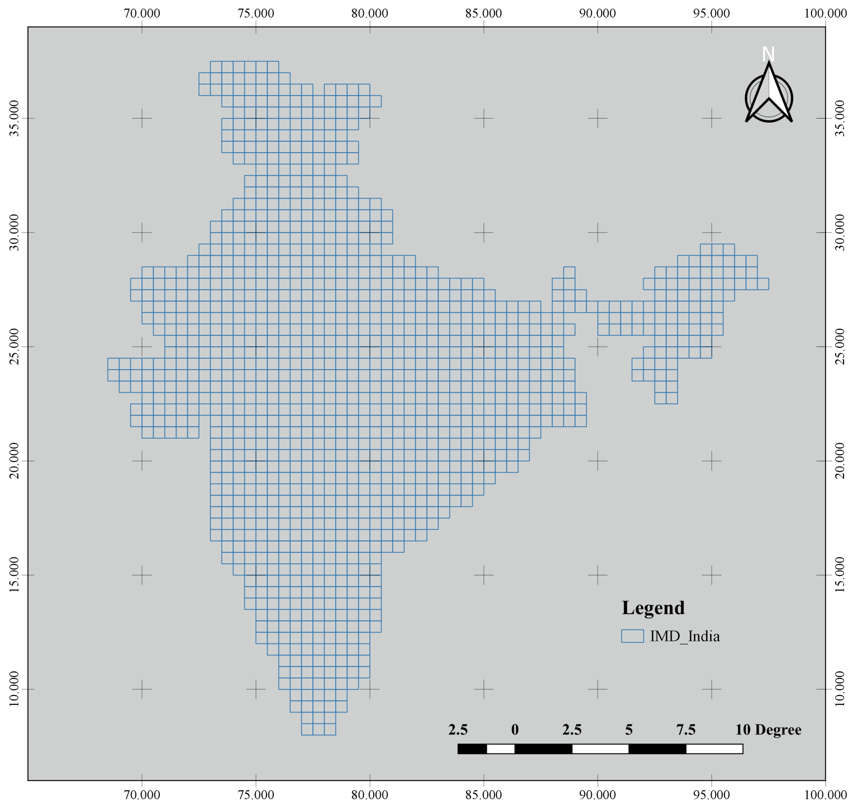


Figure 7.5: IMD Spatial Metadata Grids

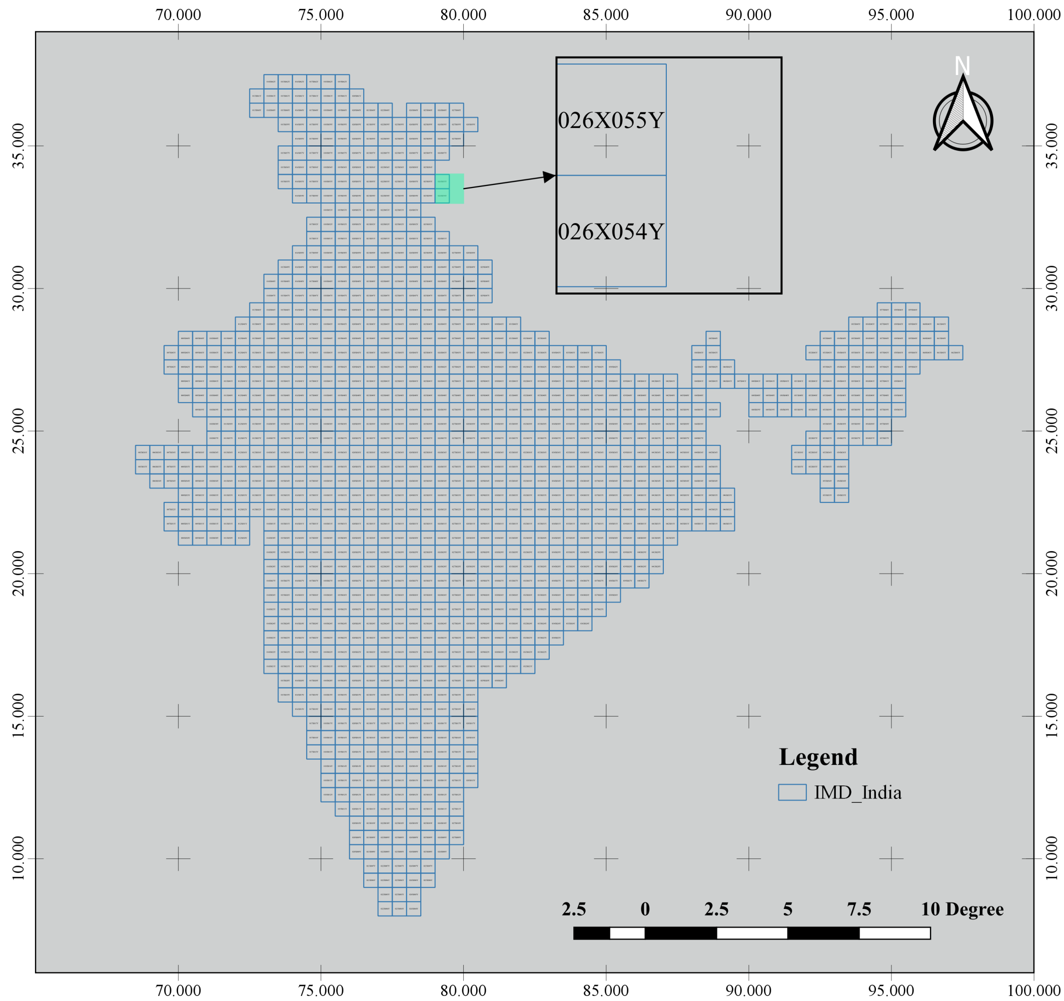


Figure 7.6: IMD Spatial Metadata Grids Map With Zoom

7.3 Pre-Processing of CanESM2 Data-sets (Predictors)

7.3.1 Generation of Spatial Metadata of CanESM2 Grids

The spatial metadata of CanESM2 data-set are generated using the longitude and latitude indices given in Table 6.3 and Table 6.4 respectively and processed using the algorithms presented in sections 6.4 of Chapter 6.

Spatial indices of the CanESM2 spatial meta-data are encoded in WKT format using the algorithm presented in section 6.4 of Chapter 06, which is shown in Figure 7.7. The map based visualization of CanESM2 spatial meta-

data is shown in Figure 7.8.

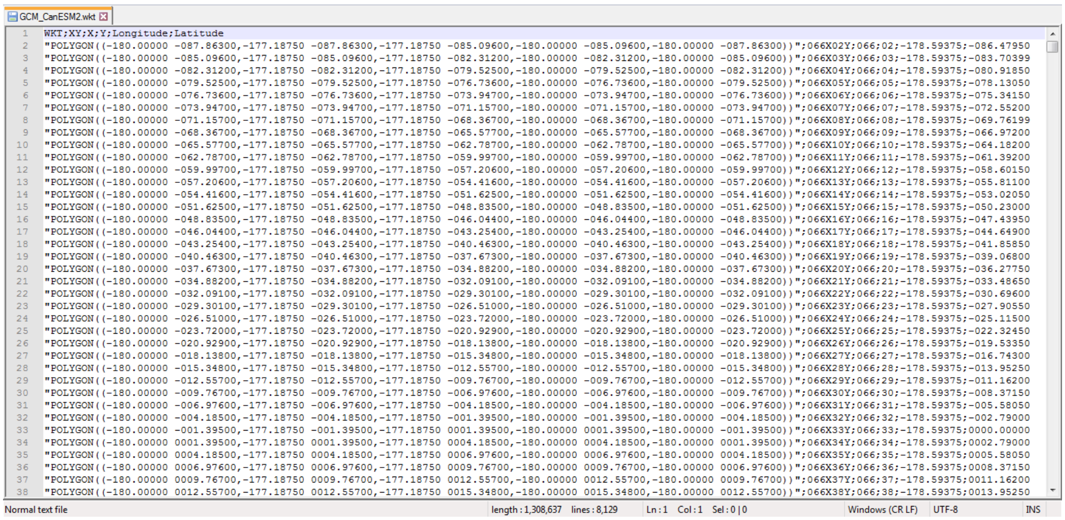


Figure 7.7: CanESM2 Spatial Metadata in WKT

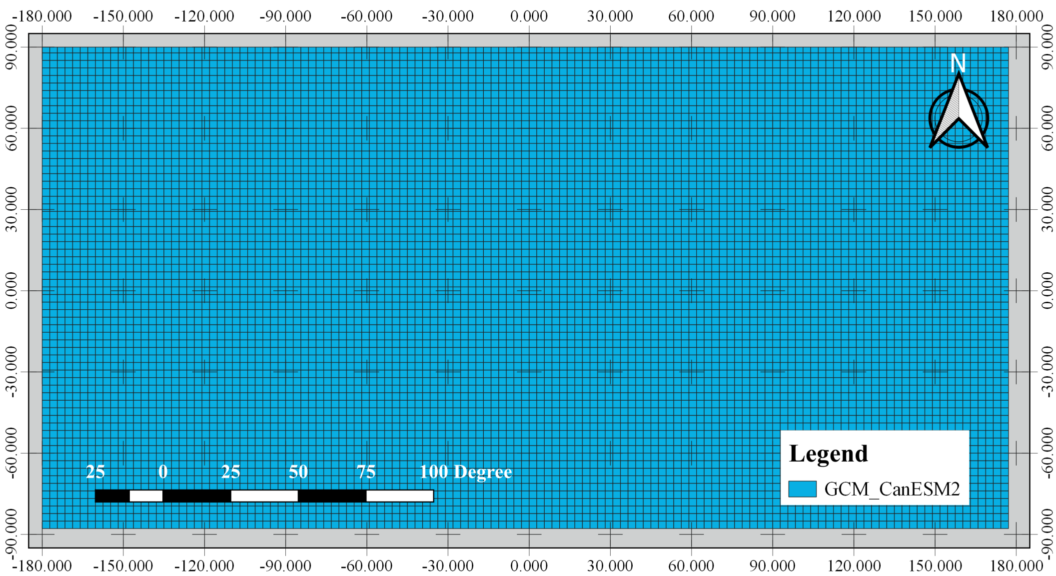


Figure 7.8: CanESM2 Grids Map

7.3.2 Generation of Metadata of CanESM2 Grids for India

Metadata of CanESM2 Grids for India in form of indices are generated using the overlay analysis of IMD and CanESM2 spatial grids using the procedure presented in section 6.4.2.2. The map based visualization of IMD and CanESM2 spatial grid layers are shown in Figure 7.9 before carrying out the

overlay analysis. Figure 7.10 depicts CanESM2 grids for India that are overlaid over IMD grids. Figure 7.11 depicts the CanESM2 grids for India in gray colour. The results of overlay analysis for CanESM2 are stored in Index file in csv file format as shown in Figure 7.12.

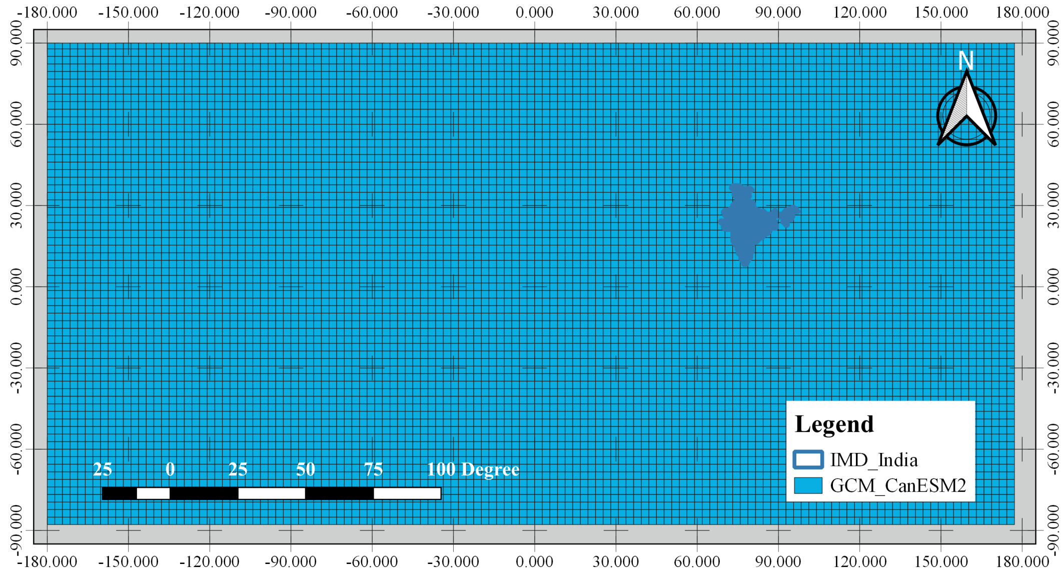


Figure 7.9: IMD and CanESM2 Grids Overlay Map

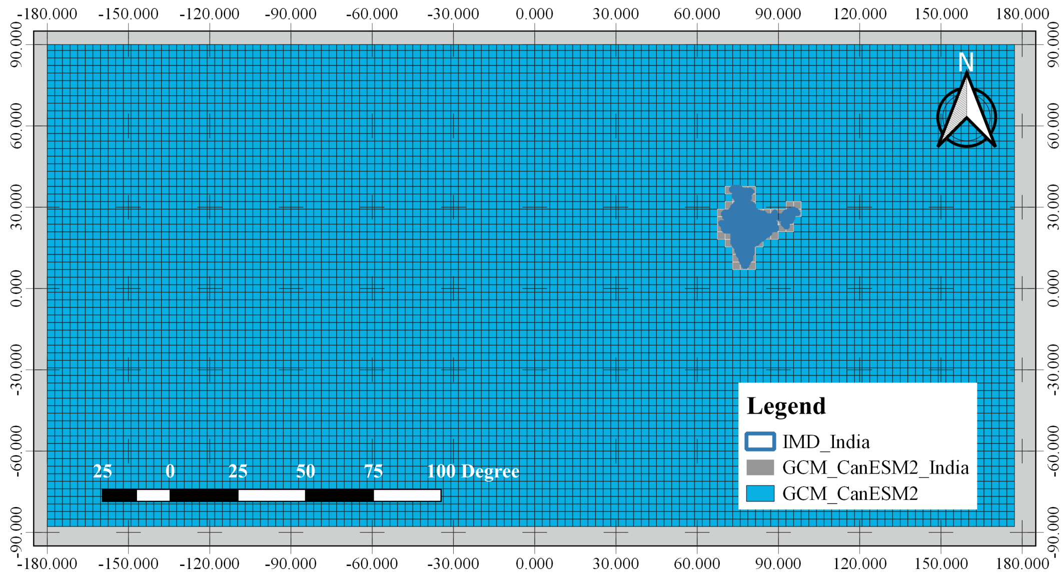


Figure 7.10: IMD Grids, CanESM2 Grids, and CanESM2 Grids for India Overlay Map

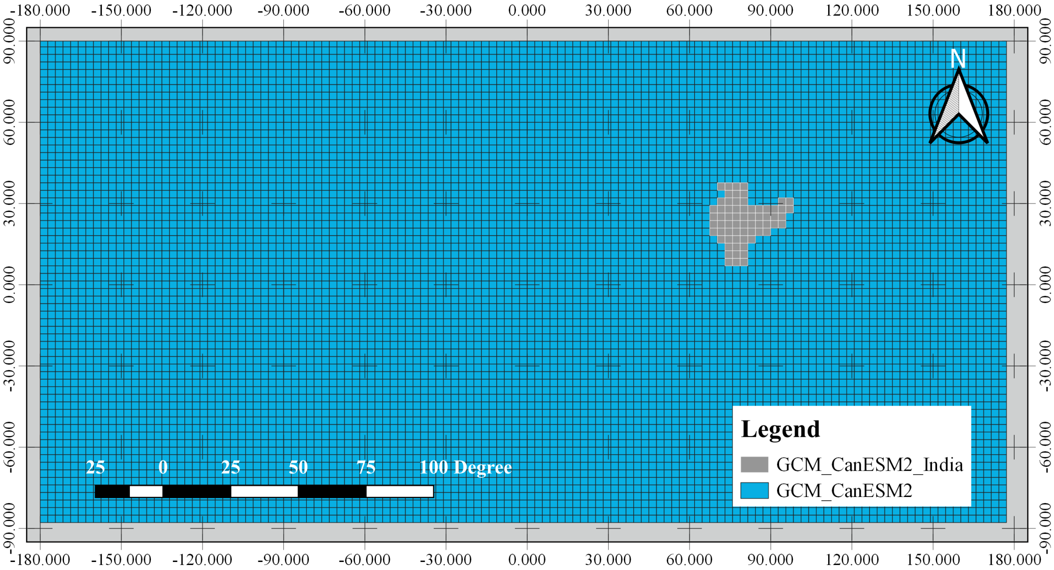


Figure 7.11: CanESM2 Grids, and CanESM2 Grids for India Overlay Map

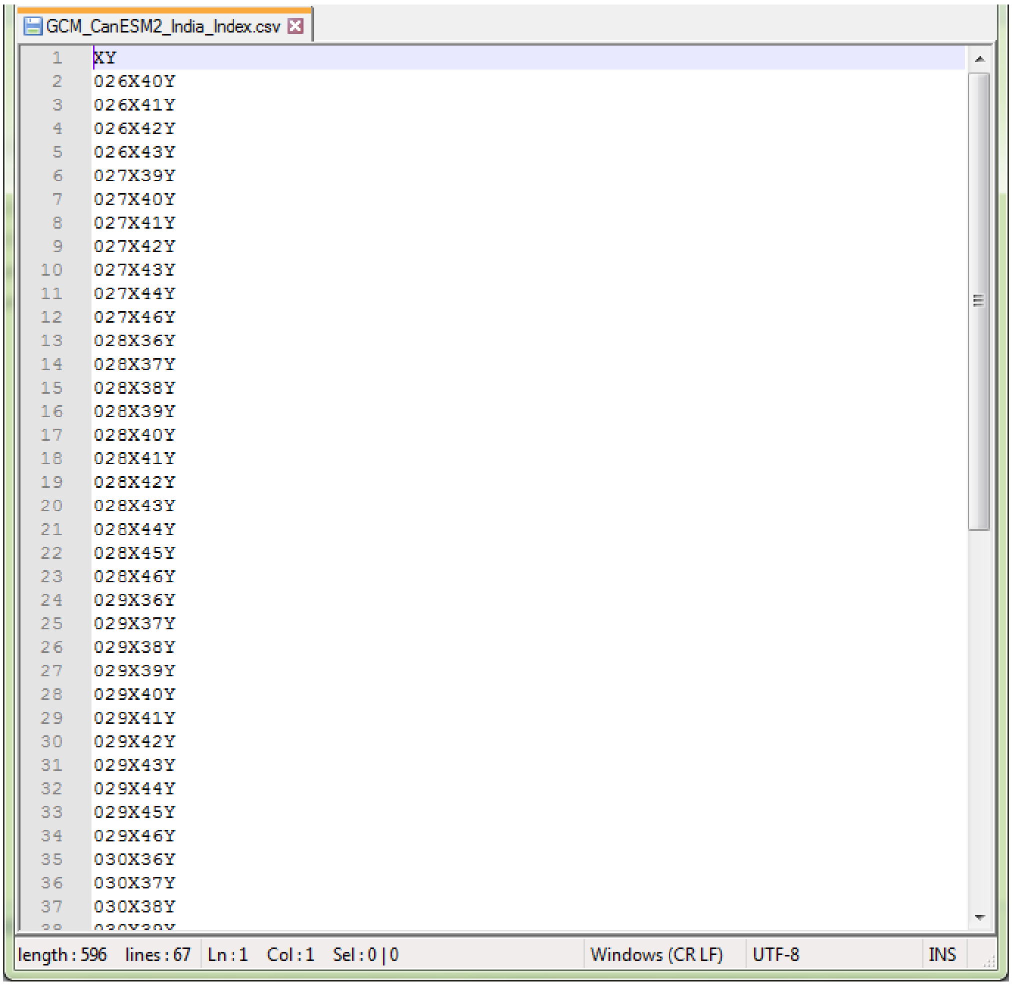
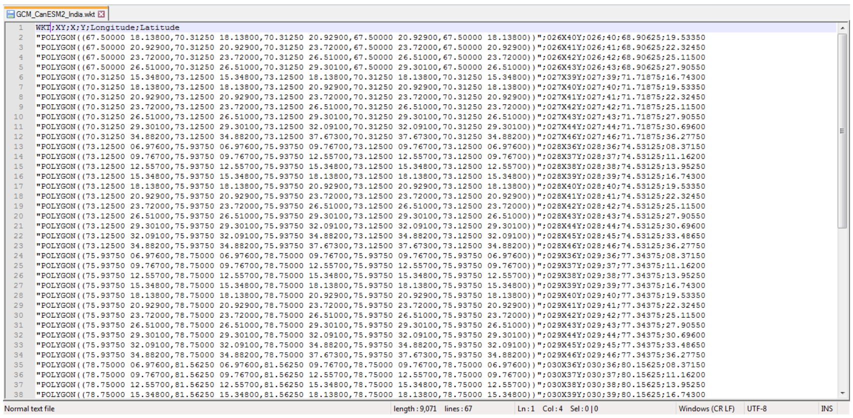


Figure 7.12: CanESM2 Indices for India

7.3.3 Extraction of Spatial Metadata for CanESM2 Grids of India

The spatial meta-data of CanESM2 for India is generated using the algorithm presented in section 6.4.2.3 using index file generated in section 6.4.2.2, which is shown in Figure 7.13. In Figure 7.14 the grids of spatial meta-data of CanESM2 for India is shown in map format. In Figure 7.15, grids are shown with their indices. IMD and CanESM2 for India spatial grid layers are juxtaposed over each other that are depicted in Figure 7.16 and Figure 7.17.



GCM_CanESM2_India.wkt

```
WKT;XY;X;Y;Longitude;Latitude
"POLYGON((67.50000 18.13800,70.31250 18.13800,70.31250 20.92900,67.50000 20.92900,67.50000 18.13800))";026X40Y;026;40;68.90625;19.53350
"POLYGON((67.50000 20.92900,70.31250 20.92900,70.31250 23.72000,67.50000 23.72000,67.50000 20.92900))";026X41Y;026;41;68.90625;22.32450
"POLYGON((67.50000 23.72000,70.31250 23.72000,70.31250 26.51000,67.50000 26.51000,67.50000 23.72000))";026X42Y;026;42;68.90625;25.11500
"POLYGON((67.50000 26.51000,70.31250 26.51000,70.31250 29.30100,67.50000 29.30100,67.50000 26.51000))";026X43Y;026;43;68.90625;27.90550
"POLYGON((70.31250 15.34800,73.12500 15.34800,73.12500 18.13800,70.31250 18.13800,70.31250 15.34800))";027X39Y;027;39;71.71875;16.74300
"POLYGON((70.31250 18.13800,73.12500 18.13800,73.12500 20.92900,70.31250 20.92900,70.31250 18.13800))";027X40Y;027;40;71.71875;19.53350
"POLYGON((70.31250 20.92900,73.12500 20.92900,73.12500 23.72000,70.31250 23.72000,70.31250 20.92900))";027X41Y;027;41;71.71875;22.32450
"POLYGON((70.31250 23.72000,73.12500 23.72000,73.12500 26.51000,70.31250 26.51000,70.31250 23.72000))";027X42Y;027;42;71.71875;25.11500
"POLYGON((70.31250 26.51000,73.12500 26.51000,73.12500 29.30100,70.31250 29.30100,70.31250 26.51000))";027X43Y;027;43;71.71875;27.90550
"POLYGON((70.31250 29.30100,73.12500 29.30100,73.12500 32.09100,70.31250 32.09100,70.31250 29.30100))";027X44Y;027;44;71.71875;30.69600
"POLYGON((70.31250 34.88200,73.12500 34.88200,73.12500 37.67300,70.31250 37.67300,70.31250 34.88200))";027X46Y;027;46;71.71875;36.27750
"POLYGON((73.12500 06.97600,75.93750 06.97600,75.93750 09.76700,73.12500 09.76700,73.12500 06.97600))";028X36Y;028;36;74.53125;08.37150
"POLYGON((73.12500 09.76700,75.93750 09.76700,75.93750 12.55700,73.12500 12.55700,73.12500 09.76700))";028X37Y;028;37;74.53125;11.16200
"POLYGON((73.12500 12.55700,75.93750 12.55700,75.93750 15.34800,73.12500 15.34800,73.12500 12.55700))";028X38Y;028;38;74.53125;13.95250
"POLYGON((73.12500 15.34800,75.93750 15.34800,75.93750 18.13800,73.12500 18.13800,73.12500 15.34800))";028X39Y;028;39;74.53125;16.74300
"POLYGON((73.12500 18.13800,75.93750 18.13800,75.93750 20.92900,73.12500 20.92900,73.12500 18.13800))";028X40Y;028;40;74.53125;19.53350
"POLYGON((73.12500 20.92900,75.93750 20.92900,75.93750 23.72000,73.12500 23.72000,73.12500 20.92900))";028X41Y;028;41;74.53125;22.32450
"POLYGON((73.12500 23.72000,75.93750 23.72000,75.93750 26.51000,73.12500 26.51000,73.12500 23.72000))";028X42Y;028;42;74.53125;25.11500
"POLYGON((73.12500 26.51000,75.93750 26.51000,75.93750 29.30100,73.12500 29.30100,73.12500 26.51000))";028X43Y;028;43;74.53125;27.90550
"POLYGON((73.12500 29.30100,75.93750 29.30100,75.93750 32.09100,73.12500 32.09100,73.12500 29.30100))";028X44Y;028;44;74.53125;30.69600
"POLYGON((73.12500 32.09100,75.93750 32.09100,75.93750 34.88200,73.12500 34.88200,73.12500 32.09100))";028X45Y;028;45;74.53125;33.48650
"POLYGON((73.12500 34.88200,75.93750 34.88200,75.93750 37.67300,73.12500 37.67300,73.12500 34.88200))";028X46Y;028;46;74.53125;36.27750
"POLYGON((75.93750 06.97600,78.75000 06.97600,78.75000 09.76700,75.93750 09.76700,75.93750 06.97600))";029X36Y;029;36;77.34375;08.37150
"POLYGON((75.93750 09.76700,78.75000 09.76700,78.75000 12.55700,75.93750 12.55700,75.93750 09.76700))";029X37Y;029;37;77.34375;11.16200
"POLYGON((75.93750 12.55700,78.75000 12.55700,78.75000 15.34800,75.93750 15.34800,75.93750 12.55700))";029X38Y;029;38;77.34375;13.95250
"POLYGON((75.93750 15.34800,78.75000 15.34800,78.75000 18.13800,75.93750 18.13800,75.93750 15.34800))";029X39Y;029;39;77.34375;16.74300
"POLYGON((75.93750 18.13800,78.75000 18.13800,78.75000 20.92900,75.93750 20.92900,75.93750 18.13800))";029X40Y;029;40;77.34375;19.53350
"POLYGON((75.93750 20.92900,78.75000 20.92900,78.75000 23.72000,75.93750 23.72000,75.93750 20.92900))";029X41Y;029;41;77.34375;22.32450
"POLYGON((75.93750 23.72000,78.75000 23.72000,78.75000 26.51000,75.93750 26.51000,75.93750 23.72000))";029X42Y;029;42;77.34375;25.11500
"POLYGON((75.93750 26.51000,78.75000 26.51000,78.75000 29.30100,75.93750 29.30100,75.93750 26.51000))";029X43Y;029;43;77.34375;27.90550
"POLYGON((75.93750 29.30100,78.75000 29.30100,78.75000 32.09100,75.93750 32.09100,75.93750 29.30100))";029X44Y;029;44;77.34375;30.69600
"POLYGON((75.93750 32.09100,78.75000 32.09100,78.75000 34.88200,75.93750 34.88200,75.93750 32.09100))";029X45Y;029;45;77.34375;33.48650
"POLYGON((75.93750 34.88200,78.75000 34.88200,78.75000 37.67300,75.93750 37.67300,75.93750 34.88200))";029X46Y;029;46;77.34375;36.27750
"POLYGON((78.75000 06.97600,81.56250 06.97600,81.56250 09.76700,78.75000 09.76700,78.75000 06.97600))";030X36Y;030;36;80.15625;08.37150
"POLYGON((78.75000 09.76700,81.56250 09.76700,81.56250 12.55700,78.75000 12.55700,78.75000 09.76700))";030X37Y;030;37;80.15625;11.16200
"POLYGON((78.75000 12.55700,81.56250 12.55700,81.56250 15.34800,78.75000 15.34800,78.75000 12.55700))";030X38Y;030;38;80.15625;13.95250
"POLYGON((78.75000 15.34800,81.56250 15.34800,81.56250 18.13800,78.75000 18.13800,78.75000 15.34800))";030X39Y;030;39;80.15625;16.74300
```

Normal text file | length : 9,071 lines : 67 | Ln : 1 Col : 4 Sel : 0 | 0 | Windows (CR LF) | UTF-8 | INS

Figure 7.13: WKT Spatial Metadata of CanESM2 for India

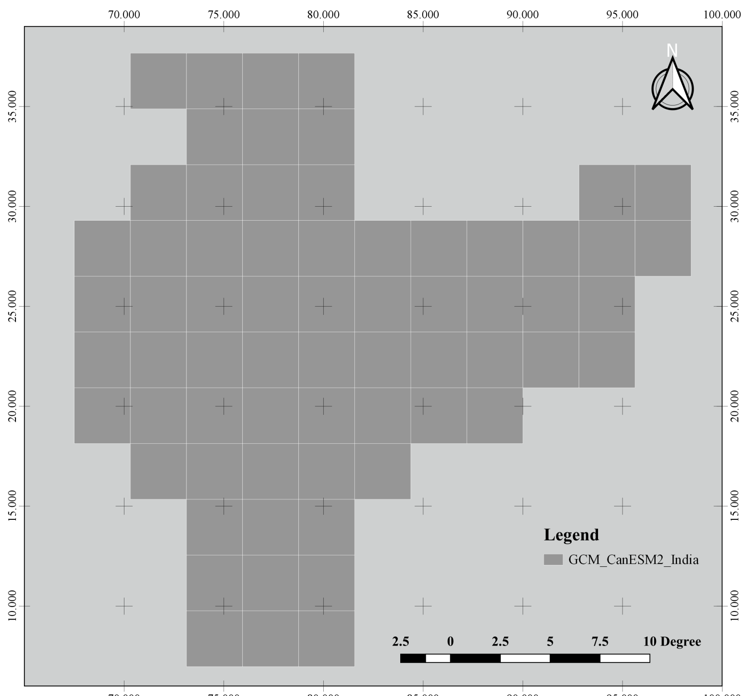


Figure 7.14: CanESM2 Grids for India Map

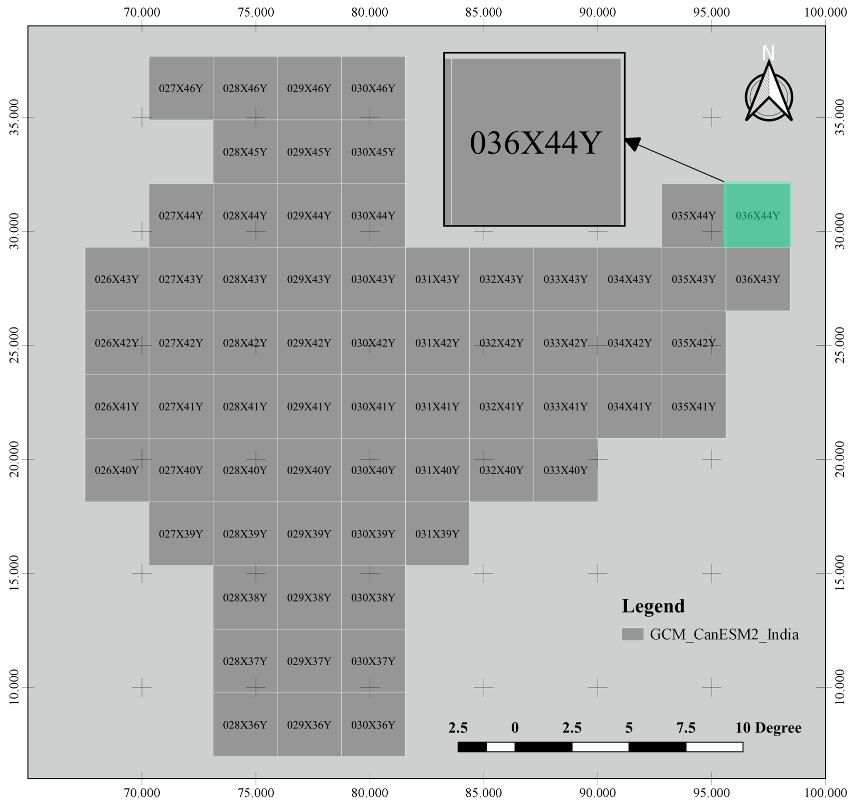


Figure 7.15: CanESM2 Grids for India Map with Zoom

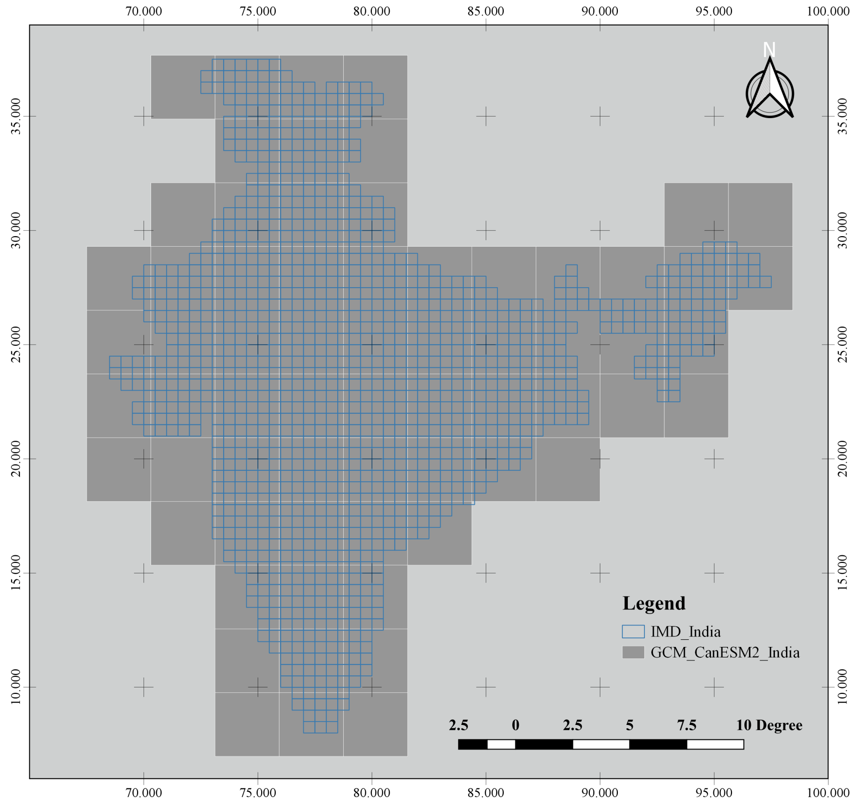


Figure 7.16: IMD Grids and CanESM2 Grids for India Overlay Map

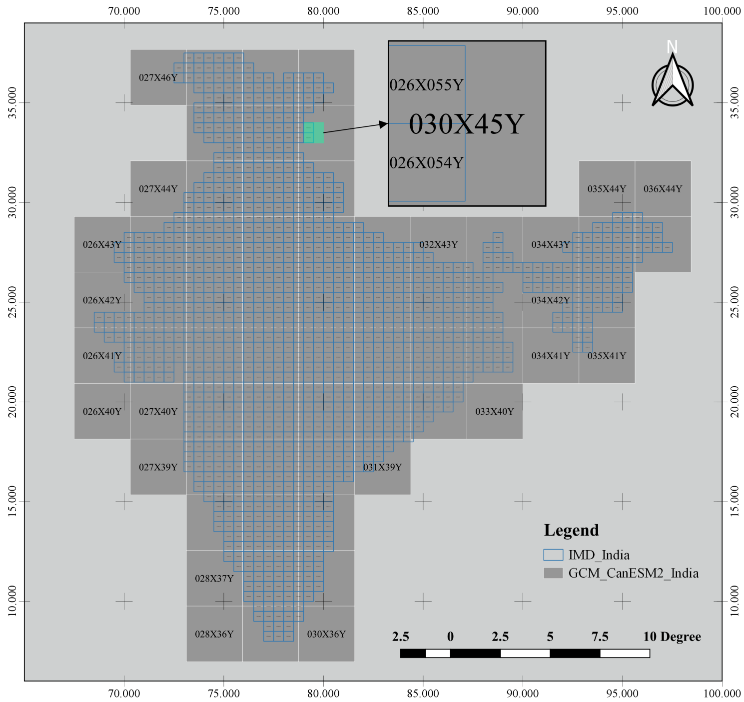


Figure 7.17: IMD Grids and CanESM2 Grids for India Zoomed Overlay Map

7.4 Spatial Analysis

The relevant CanESM2 grid corresponding to IMD grid is extracted using spatial analysis module generated by the algorithm presented in the Pseudo Code 11 as given in section 6.5. The process is depicted in Figure 7.18. The module takes four inputs (viz. data-source name, AOI name, GCM name and grid Index) and produces the CanESM2 grid index.

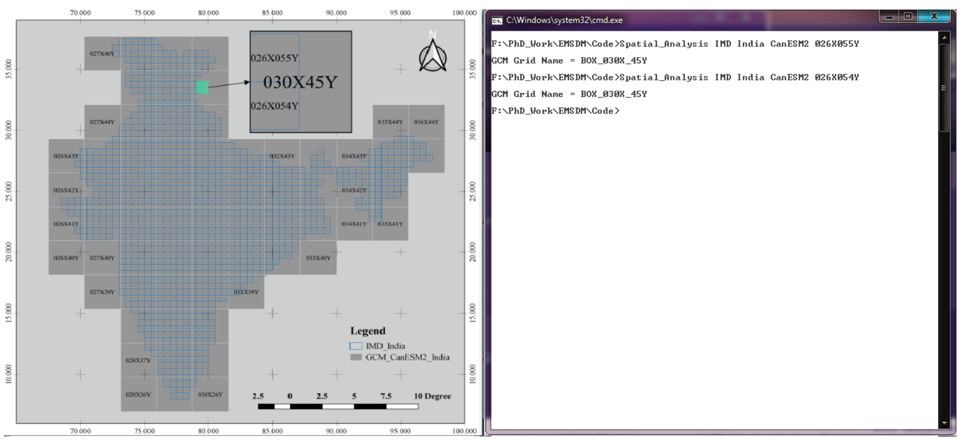


Figure 7.18: Spatial Analysis Result in Text Format

7.5 Temporal Analysis

Temporal Transformation is applied to IMD and CanESM2 data-sets to obtain temporally similar (daily, monthly, annually and decadal temporal scale) data-sets. Moreover, mapping of IMD data and CanESM2 data-sets is also carried out using the temporal analysis. Figure 7.19 depicts the programmatically generated directories structure and transformed data-sets.

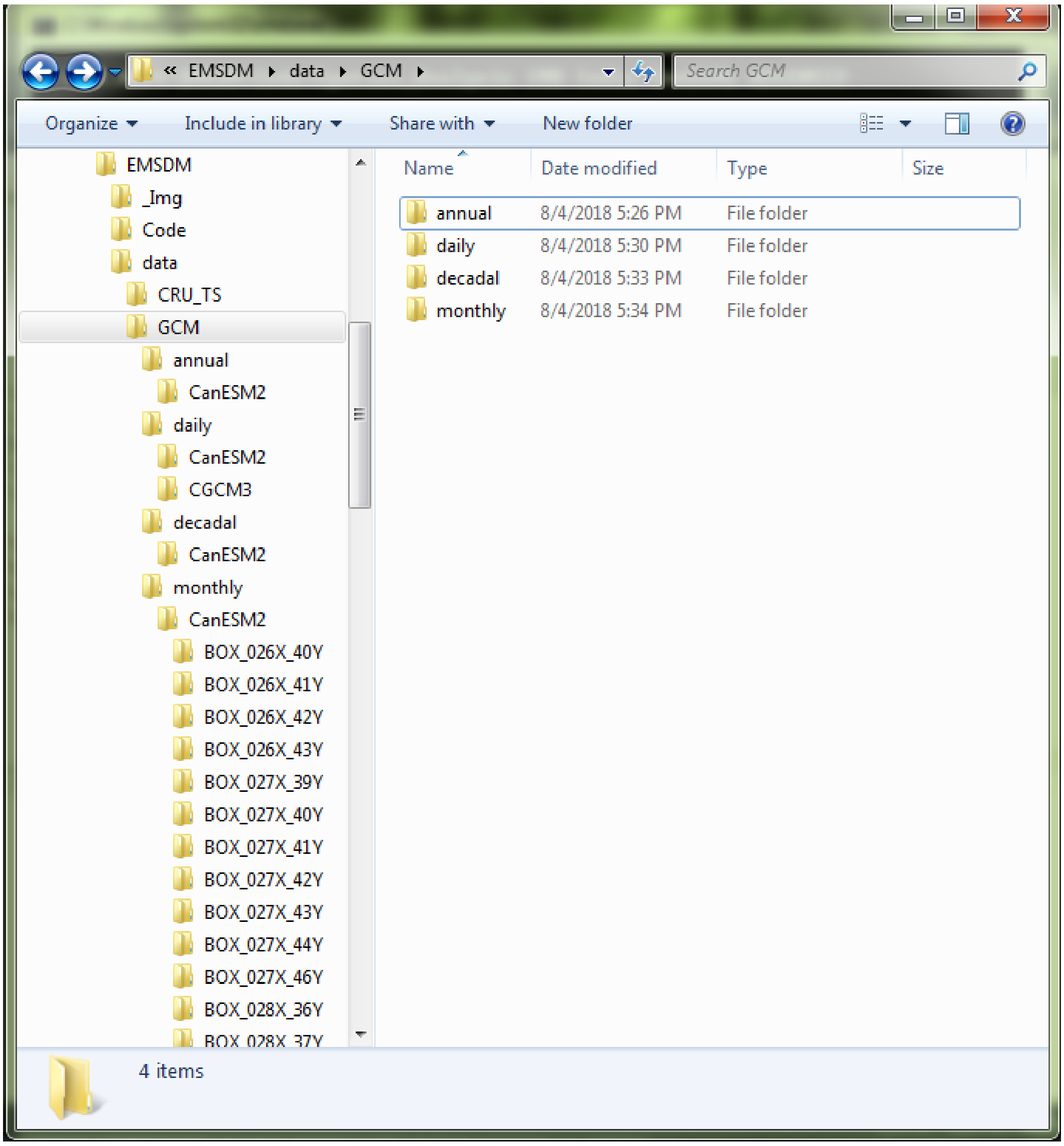


Figure 7.19: Directory Structure after Temporal Analysis

7.6 Screening Predictors

CanESM2 data-set comprises of data pertaining to twenty six parameters, which are given in Table 7.1. Screening of predictors has been carried out the using the algorithm given in Pseudo Code Listing 13 of section 6.7, so as to select the relevant predictors for the model generation.

Table 7.1: Index of Climate Variables in CanESM2 Data-set

S.No.	Parameter	Description
1	mslpgl	Mean Seal Level Pressure
2	p1_fgl	1000hPa Wind Speed
3	p1_ugl	1000hPa Zonal Wind Component
4	p1_vgl	1000hPa Meridional Wind Component
5	p1_zgl	1000hPa Relative Vorticity of Wind
6	p1thgl	1000hPa Wind Direction
7	p1zhgl	1000hPa Divergence of True Wind
8	p500gl	500hPa Geopotential
9	p5_fgl	500hPa Wind Speed
10	p5_ugl	500hPa Zonal Wind Component
11	p5_vgl	500hPa Meridional Wind Component
12	p5_zgl	500hPa Relative Vorticity of Wind
13	p5thgl	500hPa Wind Direction
14	p5zhgl	500hPa Divergence of True Wind
15	p850gl	850hPa Geopotential
16	p8_fgl	850hPa Wind Speed
17	p8_ugl	850hPa Zonal Wind Component
18	p8_vgl	850hPa Meridional Wind Component
19	p8_zgl	850hPa Relative Vorticity of Wind
20	p8thgl	850hPa Wind Direction
21	p8zhgl	850hPa Divergence of True Wind
22	prcpgl	Total Precipitation
23	s500gl	500hPa Specific Humidity
24	s850gl	850hPa Specific Humidity
25	shumgl	1000hPa Specific Humidity
26	tempgl	Air Temperature at 2m

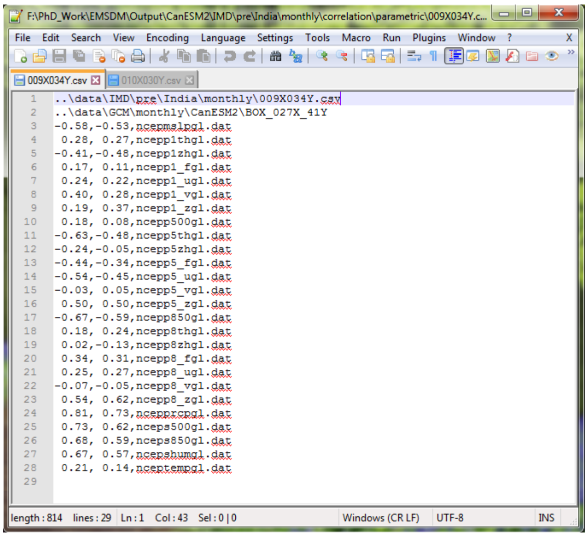


Figure 7.20: Results of Correlation Analysis for IMD Grid 009X034Y

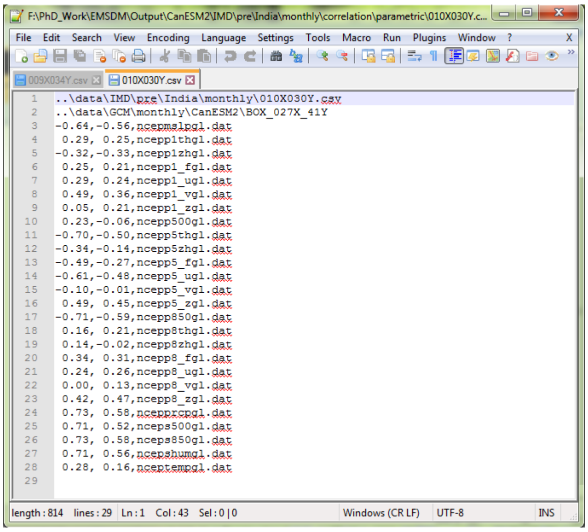


Figure 7.21: Results of Correlation Analysis for IMD Grid 010X030Y

The Figures 7.20 and 7.21 depicts the results of correlation for IMD grid 009X034Y and 010X030Y respectively corresponding to their spatially referenced CanESM2 grids. In the similar way the process is applied to the all IMD grids, and the correlations coefficients are stored in their corresponding files programmatically. The Figures 7.22 and 7.23 depicts the programming interface and the execution time to compute the Pearson and Spearman-Rho statistics. The execution time of Pearson and Spearman-Rho statistics are 1199 seconds and 624 seconds respectively.

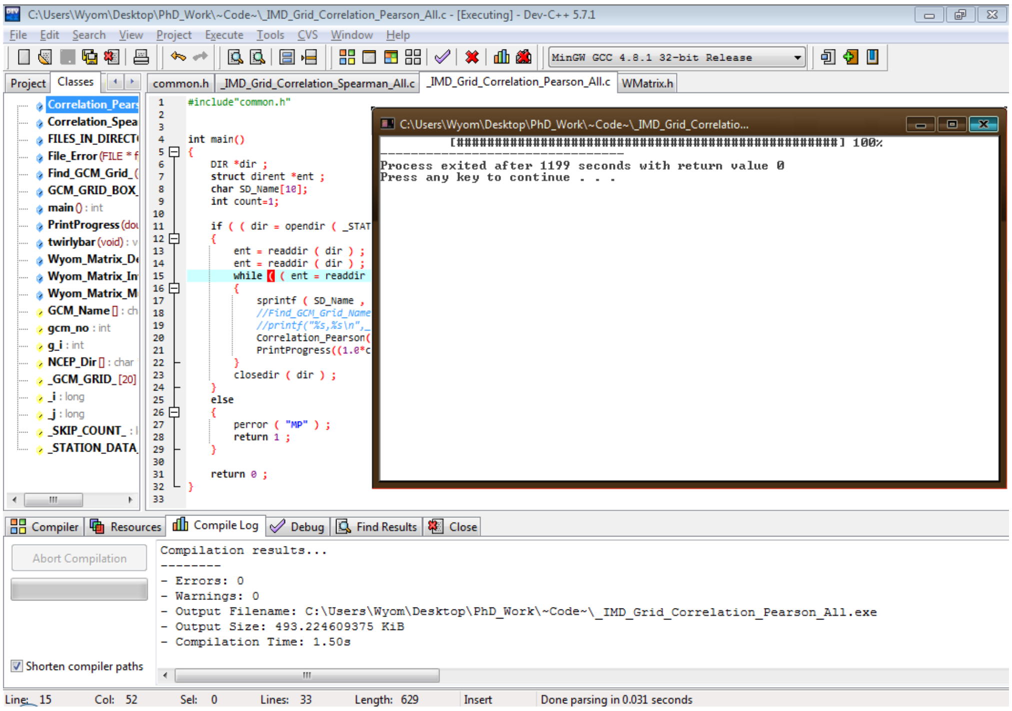


Figure 7.22: Correlation Module Execution-Time with Pearson method for Whole India

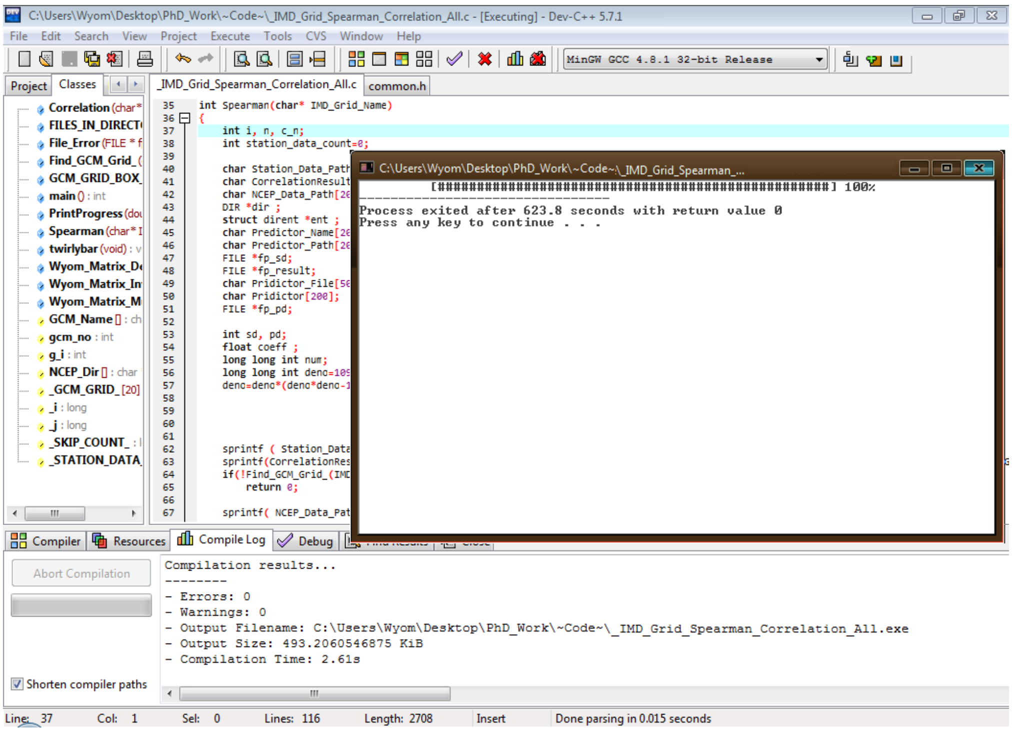
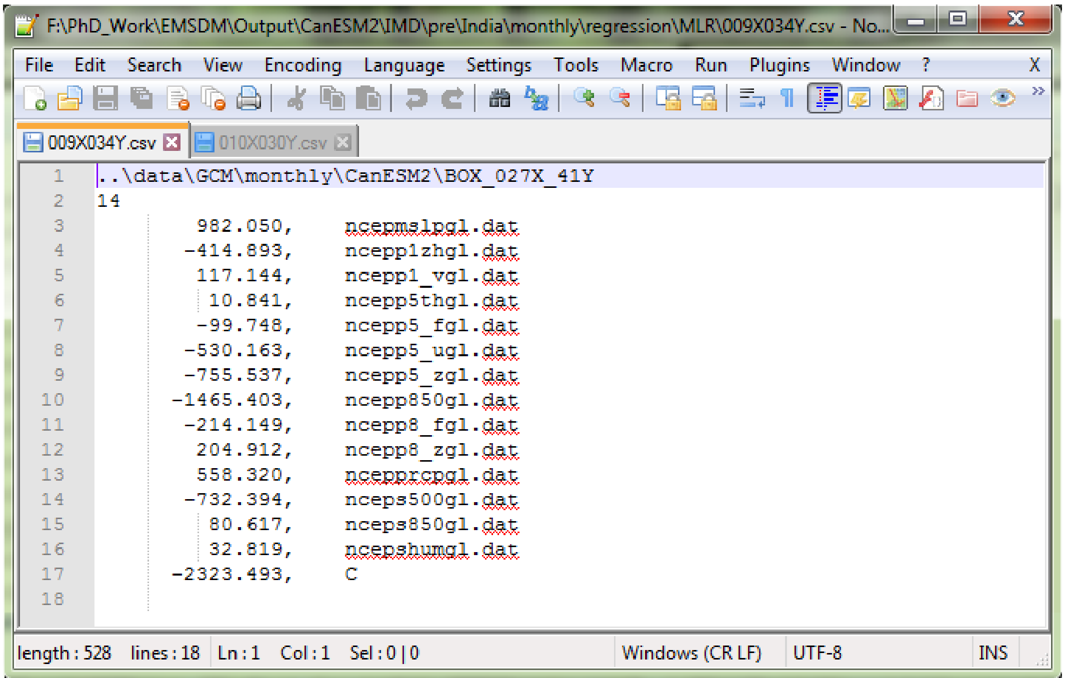


Figure 7.23: Correlation Module Execution-Time with Spearman-Rho method for Whole India

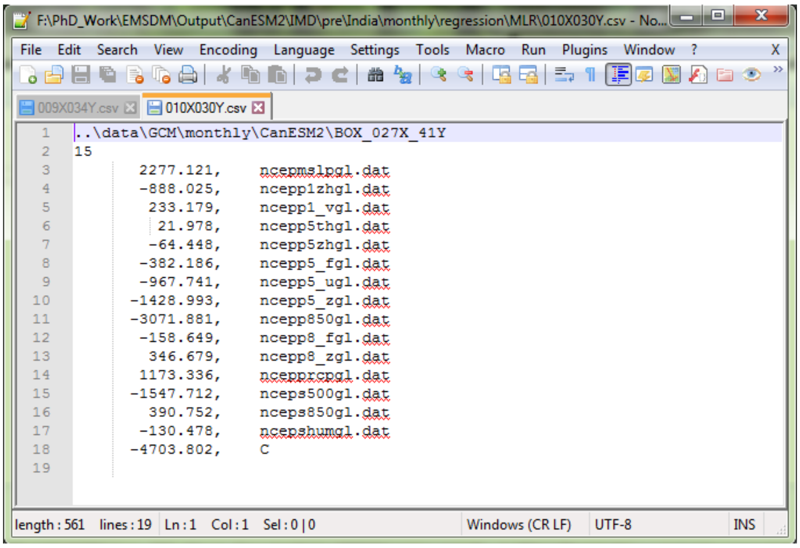
7.7 Model Generation

The statistical downscaling models for IMD grids are programmatically generated using MLR and MvLR techniques as discussed in section 6.8. EMSDM programmatically generates disparate mathematical models for each IMD grids through automating the algorithm discussed in Pseudo Code 14. The Figures 7.24 and 7.25 depicts the model parameters for IMD grid 009X034Y and 010X030Y respectively corresponding to their spatially referenced CanESM2 grids. The model parameters and their coefficients vary from grid to grid. In comparison to existing statistical downscaling tools like SDSM, EMSDM automates the process of model development for multiple grids in one execution cycle. that saves the time of the climate researchers for model generation for multiple grids.



```
..\data\GCM\monthly\CanESM2\BOX_027X_41Y
14
      982.050,    ncepmslpgl.dat
     -414.893,    ncepp1zhgl.dat
      117.144,    ncepp1_vgl.dat
       10.841,    ncepp5thgl.dat
      -99.748,    ncepp5_fgl.dat
     -530.163,    ncepp5_ugl.dat
     -755.537,    ncepp5_zgl.dat
    -1465.403,    ncepp850gl.dat
     -214.149,    ncepp8_fgl.dat
      204.912,    ncepp8_zgl.dat
      558.320,    nceppropgl.dat
     -732.394,    nceps500gl.dat
       80.617,    nceps850gl.dat
       32.819,    ncepshumgl.dat
    -2323.493,    C
```

Figure 7.24: Regression Results for IMD Grid 009X034Y



```
..\data\GCM\monthly\CanESM2\BOX_027X_41Y
15
   2277.121,    ncepmslpgl.dat
   -888.025,    ncepp1zhgl.dat
    233.179,    ncepp1_vgl.dat
     21.978,    ncepp5thgl.dat
    -64.448,    ncepp5zhgl.dat
   -382.186,    ncepp5_fgl.dat
   -967.741,    ncepp5_ugl.dat
  -1428.993,    ncepp5_zgl.dat
  -3071.881,    ncepp850gl.dat
   -158.649,    ncepp8_fgl.dat
    346.679,    ncepp8_zgl.dat
   1173.336,    nceprcpgl.dat
  -1547.712,    nceps500gl.dat
    390.752,    nceps850gl.dat
   -130.478,    ncepshumgl.dat
  -4703.802,    C
```


Figure 7.25: Regression Results for IMD Grid 010X030Y

7.8 Time Series Analysis

In order to demonstrate the applicability of the downscaling models generated using the previous process, time series is generated for the IMD grid 022X053Y having longitude of 77.25° and latitude of 32.75° . The grid position is shown in the Figure 7.27 and corresponding monthly time series of precipitation is shown in Figure 7.26.

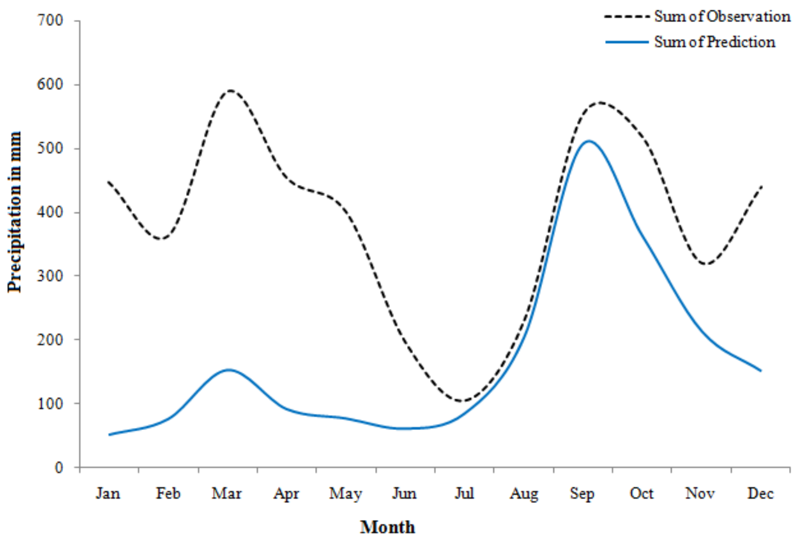


Figure 7.26: Time Series for IMD Grid 022X053Y

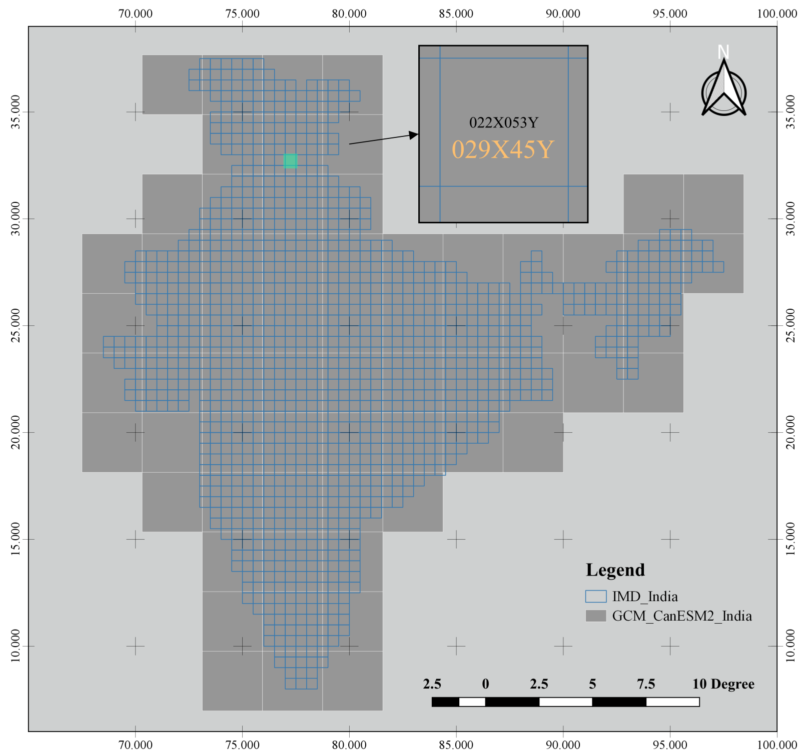


Figure 7.27: Selected IMD Grid 022X053Y for Time Series Generation

7.9 Geovisualiazation of Results

The downscaling results are available to analyst for geo-visualization through python based web-portal. Figure 7.28 depicts the downscaled information of climate variable (precipitation). In the same figure context menu provide the information about location, index and decadal precipitation as obtained from EMSDM for different decades. Figure 7.29 depicts the overlayed downscaled information of climate variable (precipitation) for India . In the same figure context menu provide the information about location, index and decadal precipitation as obtained from EMSDM for different decades.Analyst can query information about the grids as shown in figure. Through the web-based portal, downscaling results generated by EMSDM is widely available to the decision makers.

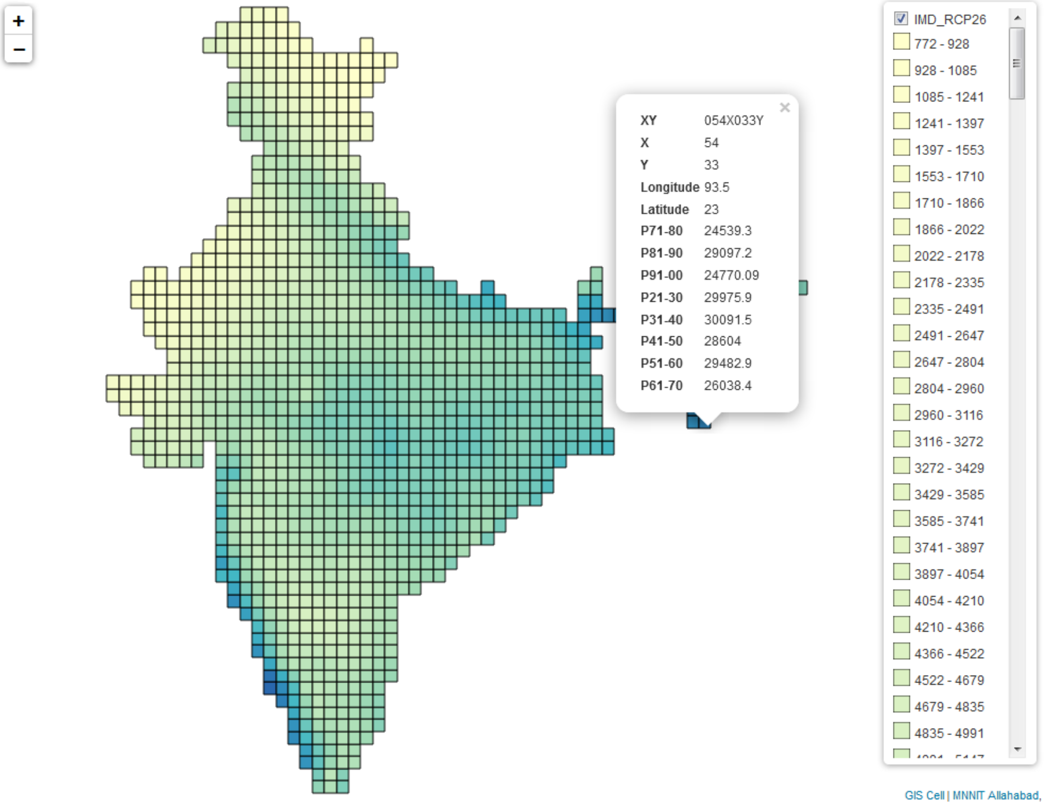


Figure 7.28: Geo-visualization of EMSDM for India

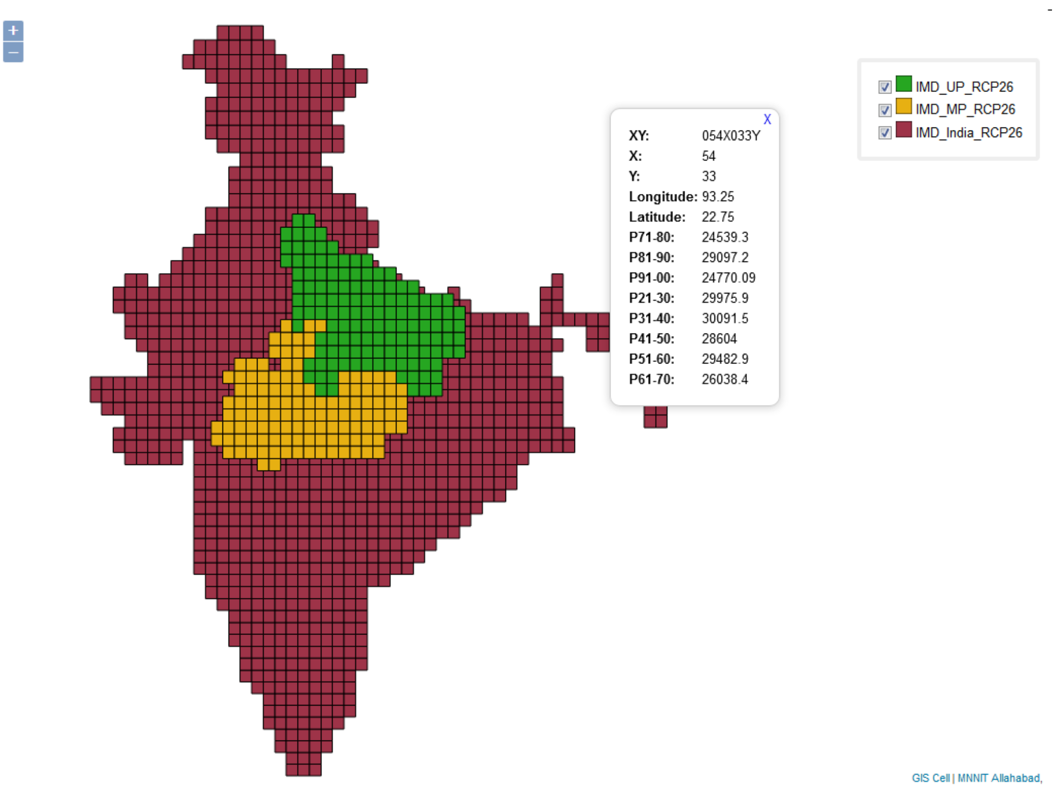


Figure 7.29: Geo-visualization EMSDM Result for India, UP and MP

7.10 Concluding Remarks

In this chapter the application of the EMSDM is being carried out using CanESM2 GCM data-set and IMD data-set for India In order to demonstrate this applicability for successfully downscaling the Climate Data-set. EMSDM automates the different underlying processes and generates valuable structured spatial meta-data for GCM and local Climate data-set, that can be efficiently utilized for statistical downscaling of climate data for specified area of interest.

down procedures required for implementing the EMSDM has been discussed in detail. Algorithms and procedures of EMSDM are programmed in C language. Henceforth EMSDM can be implemented irrespective of the Operating Systems viz. Windows, Unix, Linux, macOS etc and possess advantage of interoperability over widely used Window Based SDSM software. Moreover, in addition of carrying out statistical downscaling, EMSDM generates the interoperable spatial data for corresponding GCM and local grids. These spatial data can be used for different climate studies. After discussion on implementation of EMSDM in this chapter, in Chapter 7, applicability of EMSDM is demonstrated using CanESM2 as GCM data-set, IMD as local data-set and India as a selected AOI.

Chapter 8

Conclusions and Recommendations

8.1 Conclusions

Following conclusions have been drawn in the present research work:

1. Statistical downscaling is most widely applied method for downscaling the climate data. Most widely used software like Statistical DownScaling Model (SDSM) and Long Ashton Research Station Weather Generator (LARS-WG) implement statistical downscaling method for downscaling the climate model at local scale. In these software, execution time of the downscaling process for single grid is approximately 30 minutes. In addition, these softwares are very time inefficient for multi-site Downscaling. In this thesis the developed statistical downscaling (Implemented in C language) is very time efficient for one grid at a time as well as multi-site downscaling.

2. Existing software do not provide functionality for carrying out statistical downscaling of large regions in a single execution. Henceforth, using the presently available software, significant time and manual efforts are required to carry to statistical downscaling in a piecewise manner. In this thesis, proposed model namely EMSDM is able to automate multi-site downscaling efficiently. Hence, human intervention is not required

for multi-site downscaling using EMSDM.

3. Existing software do not provide functionality for carrying out statistical downscaling of large regions in a single execution. Henceforth, using the presently available software, significant time and manual efforts are required to carry to statistical downscaling in a piecewise manner.

4. On basis of review of existing research works, it can be concluded that generalized computational downscaling model for the application in given specified AOI irrespective of its geographical extent is not available.

5. In comparison to statistical downscaling techniques, dynamical downscaling techniques are still computationally intensive in terms of computational time and also entail a considerable amount of expensive hardware. Henceforth, statistical downscaling techniques make them appropriate for uncertainty studies since it becomes feasible to perform downscaling for various types of climate scenarios.

6. Monthly precipitation or temperature data are more suitable to develop the downscaling models since monthly downscaling models are usually more appropriate than the daily models in context of the predictive capabilities.

7. The salient features of EMSDM are:

    (a) Downscaling for given area of interest using different types of climate data like precipitation, temperature can be carried out.

    (b) It is possible to downscale the data irrespective of geographical extent.

    (c) It is scalable. It is possible to extend the framework to included new computational algorithm(s) for model development.

(d) It comprises of Web GIS component for effective geovisualization and querying of downscaled data in form of vector maps. Henceforth, downscaling results are available to decision makers for further investigation.

(e) It implements the parametric as well as non-parametric approaches for modal parameter selection. Henceforth, it provides the flexibility to filter out extreme events for model development if required by decision maker.

(f) In comparison of the other approaches for carrying out downscaling for large regions like country, the execution time for carrying out downscaling for large regions is significantly less.

(g) It provide the functionality to the generate spatial metadata grids for GCM and local observations. These spatial meta data data-sets can be assist investigators to spatio-temporally analyse the distribution of climatological parameters like temperature, precipitation etc. and further facilitate geo-statistical analysis of these parameters.

(h) It can be used to carry out sensitivity analysis of GCM parameters. Analysts can provide their set of GCM outputs with specified order of their weighing to develop regression model and subsequently generate time series to analyse the effect of these parameters on climate prediction.

8.2 Future Recommendatation

Based on the EMSDM, a full-scale standalone cloud based system may be developed and deployed so that users can utilize the system for further investigation of climate change in specified area of interest.

Bibliography

Abatzoglou, J. T. and Brown, T. J. (2012). A comparison of statistical downscaling methods suited for wildfire applications. *International Journal of Climatology*, 32(5):772–780.

Alexander, D. (2012). *Bringing the Sun Down to Earth*, volume 5.

Allaby, M. (2009). *Atmosphere : A Scientific History of Air, Weather and Climate*. Facts on File.

AM, R., P, E., A, G., and J, M. (2017). Spatial Downscaling Using Bias Correction Approach.

Ashiq, M. W., Zhao, C., Ni, J., and Akhtar, M. (2010). GIS-based high-resolution spatial interpolation of precipitation in mountain-plain areas of Upper Pakistan for regional climate change impact studies. *Theoretical and Applied Climatology*, 99(3-4):239–253.

Banakar, V. K., Baidya, S., Piotrowski, A. M., and Shankar, D. (2017). Indian summer monsoon forcing on the deglacial polar cold reversals. *Journal of Earth System Science*, 126(6):1–11.

Barsugli, J., Nowak, K., Rajagopalan, B., Prairie, J. R., and Harding, B. (2009). Comment on "when will Lake Mead go dry?" by T. P. Barnett and D. W. Pierce. *Water Resources Research*, 45(9):1–6.

Bauer, S. J. and Lammer, H. (2004). *Planetary Aeronomy Atmosphere : Atmosphere Environments in Planetary Systems*. Springer-Verlag.

Benestad, R. E., https://doi.org/10.1142/6908eliang Chen, DBenestad, R. E., Chen, D., Hanssen-bauer, I., Trzaska, S., Schnarr, E., Wikipedia, ... Hegewisch, K. . M. A. C. A. . M. . D. M., Hanssen-bauer, I., Trzaska, S., Schnarr, E., Wikipedia, Wilby, R. L., Wigley, T. M., and Hegewisch, K. (2018). Multivariate Adaptive Constructed Analogs (MACA) Downscaling Method.

Bera, K. (2016). Climate Change Impact on Land Use / Land Covers Change Through Remote Sensing , A Case Study of Dangra Sub- watershed , Bankura and Puruliya District. *International Journal of Satellite Communication & Remote Sensing*, 2(1):1–6.

Bera, S. (2017). Trend Analysis of Rainfall in Ganga Basin , India during 1901-2000. *American Journal of Climate Change*, 6:116–131.

Bhuvandas, N., Timbadiya, P. V., Patel, P. L., and Porey, P. D. (2014). Review of downscaling methods in climate change and their role in hydrological studies. *International Journal of Environmental, Ecological, Geological and Marine Engineering*, 8(10):648–653.

Biesbroek, G. R., Swart, R. J., Carter, T. R., Cowan, C., Henrichs, T., Mela, H., Morecroft, M. D., and Rey, D. (2010). Europe adapts to climate change: Comparing National Adaptation Strategies. *Global Environmental Change*, 20(3):440–450.

Blocker, A. B. (1982). Winter Weather and Cardiovascular Mortality. 72(3):261–265.

Boe, J. and Terray, L. (2008). A weather-type approach to analyzing winter precipitation in France: Twentieth-century trends and the role of anthropogenic forcing. *Journal of Climate*, 21(13):3118–3133.

Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., Kushnir, Y., Kimoto, M., Meehl, G. A., Msadek, R., Mueller, W. A., Taylor, K. E., Zwiers, F., Rixen, M., Ruprich-Robert, Y., and Eade, R. (2016). The Decadal Climate Prediction Project (DCPP) contribution to CMIP6. *Geoscientific Model Development*, 9(10):3751–3777.

Bordes, A. (2010). *New algorithms for large-scale support vector machines*. PhD thesis.

Bostrom, A., Morgan, M. G., Fischhoff, B., and Read, D. (1994). What Do People Know About Global Climate Change? 1. Mental Models. *Risk Analysis*, 14(6):959–970.

Both, C., Artemyev, A. V., Blaauw, B., Cowie, R. J., Dekhuijzen, A. J., Eeva, T., Enemar, A., Gustafsson, L., Ivankina, E. V., Jarvinen, A., Metcalfe, N. B., Nyholm, N. E. I., Potti, J., Ravussin, P.-A., Sanz, J. J., Silverin, B., Slater, F. M., Sokolov, L. V., Torok, J., Winkel, W., Wright, J., Zang, H., and Visser, M. E. (2004). Large-scale geographical variation confirms that climate change causes birds to lay earlier. *Proceedings of the Royal Society B: Biological Sciences*, 271(1549):1657–1662.

Buzan, B. (1985). After hegemony: cooperation and discord in the world political economy. *International Affairs*, 61(2):290–291.

Bychkov, V. L., Golubkov, G. V., and Nikitin, A. I. (2010). *The Atmosphere and Ionoshere : Dynamics, Processes and Monitoring*. Springer.

Campbell, C. and Ying, Y. (2011). *Learning with Support Vector Machines*.

Cannon, A. J., Sobie, S. R., and Murdock, T. Q. (2015). Bias correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes? *Journal of Climate*, 28(17):6938–6959.

Carey, J. C., Tang, J., Templer, P. H., Kroeger, K. D., Crowther, T. W., Burton, A. J., Dukes, J. S., Emmett, B., Frey, S. D., Heskel, M. A., Jiang, L., Machmuller, M. B., Mohan, J., Panetta, A. M., Reich, P. B., Reinsch, S., Wang, X., Allison, S. D., Bamminger, C., Bridgham, S., Collins, S. L., de Dato, G., Eddy, W. C., Enquist, B. J., Estiarte, M., Harte, J., Henderson, A., Johnson, B. R., Larsen, K. S., Luo, Y., Marhan, S., Melillo, J. M., Peñuelas, J., Pfeifer-Meister, L., Poll, C., Rastetter, E., Reinmann, A. B., Reynolds, L. L., Schmidt, I. K., Shaver, G. R., Strong, A. L., Suseela, V., and Tietema, A. (2016). Temperature response of soil respiration largely unaltered with experimental warming. *Proceedings of the National Academy of Sciences*, 113(48):13797–13802.

Cavalcanti, I. F. A., Nunes, L. H., Marengo, J. A., Gomes, J. L., Silveira, V. P., and Castellano, M. S. (2017). Projections of Precipitation Changes in Two Vulnerable Regions of São Paulo State, Brazil. *American Journal of Climate Change*, 06(02):268–293.

CCCma (2018). CanESM2 predictors : CMIP5 experiments Structure of Predictor Files.

Chandler, R. and Wheater, H. (2002). Analysis of rainfall variability using generalized linear models: A case study from the west of Ireland. *Water Resources Research*, 38(10):674– 689.

Chen, H., Xu, C.-y., and Guo, S. (2012). Comparison and evaluation of multiple GCMs , statistical downscaling and hydrological models in the study of climate change impacts on runoff. *Journal of Hydrology*, 434-435:36–45.

Chen, J., Brissette, F. P., Chaumont, D., and Braun, M. (2013). Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America. *Water Resources Research*, 49(7):4187–4205.

Chen, J., Brissette, F. P., and Leconte, R. (2011). Uncertainty of downscaling method in quantifying the impact of climate change on hydrology. *Journal of Hydrology*, 401(3-4):190–202.

Chen, J., Zhang, X. J., and Li, X. (2018). A Weather Generator-Based Statistical Downscaling Tool for Site-Specific Assessment of Climate Change Impacts. *Transactions of the ASABE 2018*, 61(3):977–993.

Chen, W.-Y. (2012). *Handbook of Climate Change Mitigation.*

Chisanga, C. B., Phiri, E., and Chinene, V. R. N. (2017). Statistical Downscaling of Precipitation and Temperature Using Long Ashton Research Station

Weather Generator in Zambia : A Case of Mount Makulu Agriculture Research Station. *American Journal of Climate Change*, 6:487–512.

Chithra, N. and Thampi, S. G. (2015). Detection and Attribution of Climate Change Signals in Precipitation in the Chaliyar River Basin, Kerala, India. *Aquatic Procedia*, 4(Icwrcoe):755–763.

Christensen, J. H., Carter, T. R., Rummukainen, M., and Amanatidis, G. (2007). Evaluating the performance and utility of regional climate models: The PRUDENCE project. *Climatic Change*, 81:1–6.

Climate Research Unit (2015). Climatic Research Unit.

Committee, C., Webb, M., Bretherton, C., Douville, H., Kay, J., Klein, S., Siebesma, P., Stevens, B., Tselioudis, G., Watanabe, M., Andrews, T., Chadwick, R., Douville, H., and Pmc, C. (2014). Overview of proposed CFMIP3 / CMIP6 GCM Experiments CFMIP-2 / CMIP5 Experiment Hierarchy. (July).

Correlation, K. R. (1994). Appendix : Mann-Kendall Trend Tests. pages 1–4.

Crowley, T. J. (2007). Causes of Climate Change over the past 1000 Years Causes of Climate Change Over the Past 1000 Years. 289(5477):270–277.

Cunningham, W. P. and Cunningham, M. A. (2010). *Principles of Environmental Science - Inquiry and Application*, volume 6. McGraw-Hill Higher Education.

Curry, J. (2011). Nullifying the climate null hypothesis. *Wiley Interdisciplinary Reviews: Climate Change*, 2(6):919–924.

Diaz-Nieto, J. and Wilby, R. L. (2005). A comparison of statistical downscaling and climate change factor methods: Impacts on low flows in the River Thames, United Kingdom. *Climatic Change*, 69(2-3):245–268.

Dorman, L. I. (2004). *Cosmic Rays in the Earth Atmosphere and Underground.* Springer New York.

Duhan, D. and Pandey, A. (2013). Statistical analysis of long term spatial and temporal trends of precipitation during 1901-2002 at Madhya Pradesh, India. *Atmospheric Research*, 122:136–149.

Durack, P. J. and Taylor, K. E. (2017). CMIP6 Forcing Datasets Summary. Technical report.

Easterbrook, S. (2010). Climate change: a grand software challenge. *... of the FSE/SDP workshop on Future of software ...*, pages 99–103.

Ekström, M., Grose, M. R., and Whetton, P. H. (2015). An appraisal of downscaling methods used in climate change research. *Wiley Interdisciplinary Reviews: Climate Change*, 6(3):301–319.

Evans, B. T. (1996). Effects on trend statistics of the use of multiplicative noise for disclosure limitation. *Bureau ofthe Census.*

Eyring, V. (2016). Coupled Model Intercomparison Project Phase 6 (CMIP6) - an Overview of Activities. Technical Report November.

Eyring, V., Bony, S., Meehl, G. A., Senior, C., Stevens, B., Stouffer, R. J., and Taylor, K. E. (2015). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organisation. *Geoscientific Model Development Discussions*, 8(12):10539–10583.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, 9(5):1937–1958.

Fealy, R. and Sweeney, J. (2007). Statistical downscaling of precipitation for a selection of sites in Ireland employing a generalised linear modelling approach. *International Journal of Climatology*, 4(May):2083–2094.

Feichter, J. and Stier, P. (2012). Assessment of black carbon radiative effects in climate models. *Wiley Interdisciplinary Reviews: Climate Change*, 3(4):359–370.

Fernandez, M., Piccolo, L. S. G., Maynard, D., Wippoo, M., Meili, C., and Alani, H. (2016). Talking climate change via social media: communication, engagement and behaviour. *Proceedings of the 8th ACM Conference on Web Science*, pages 85–94.

Flato, G. M. (2011). Earth system models: An overview. *Wiley Interdisciplinary Reviews: Climate Change*, 2(6):783–800.

Fowler, H. J., Blenkinsop, S., and Tebaldi, C. (2007). Linking climate change modelling to impacts studies: Recent advances in downscaling techniques for hydrological modelling.

Franzke, C. L., O'Kane, T. J., Berner, J., Williams, P. D., and Lucarini, V. (2015). Stochastic climate theory and modeling. *Wiley Interdisciplinary Reviews: Climate Change*, 6(1):63–78.

Fuller, A., Dawson, T., Helmuth, B., Hetem, R. S., Mitchell, D., and Maloney, S. K. (2010). Physiological Mechanisms in Coping with Climate Change. *Report*, 83(5):713–720.

Fung, F., Lopez, A., and New, M. (2011). *Modelling the Impact of Climate Change on Water Resources (Google eBook)*. Blackwell Publishing Ltd, John Wiley & Sons.

Gao, J. and Sheshukov, A. (2015). Uncertainty Estimation of Hydrological Impacts of Bias-Corrected CMIP5 Climate Change Projections. Technical report.

García, L., Matthews, J., Rodriguez, D., Wijnen, M., DiFrancesco, K., and Ray, P. (2014). Beyond Downscaling: A Bottom-Up Approach to Climate Adaptation for Water Resources Management. *AGWA Report 01*, page 62.

Geerts B. and Linacre E. (1998). What are general circulation models?

George, J., Janaki, L., and Parameswaran Gomathy, J. (2016). Statistical Downscaling Using Local Polynomial Regression for Rainfall Predictions – A Case Study. *Water Resources Management*, 30(1):183–193.

Gergel, D. R., Nijssen, B., Abatzoglou, J. T., Lettenmaier, D. P., and Stumbaugh, M. R. (2017). Effects of climate change on snowpack and fire potential in the western USA. *Climatic Change*, 141(2):287–299.

Ghosh, S. and Mujumdar, P. P. (2008). Statistical downscaling of GCM simulations to streamflow using relevance vector machine. *Advances in Water Resources*, 31(1):132–146.

Giorgi, F. (1990). Simulation of Regional Climate Using a Limited Area Model Nested in a General Circulation Model. *Journal of Climate*, 3(9):941–963.

Giorgi, F. and Bates, G. T. (1989). The Climatological Skill of a Regional Model over Complex Terrain. *Monthly Weather Review*, 117(11):2325–2347.

Giorgi, F., Brodeur, C. S., and Bates, G. T. (1994). Regional climate change scenarios over the United States produced with a nested regional climate model. *Journal of Climate*, 7(3):375–399.

Giorgi, F., Hewitson, B., Christensen, J., Hulme, M., von Storch, H., Whetton, P., Jones, R., Mearns, L., and Fu, C. (2001). Regional Climate Information - Evaluation and Projections. In *Geological and Atmospheric Sciences*, chapter Chapter 10, pages 583–638.

Giorgi, F., Marinucci, M., and Visconti, G. (1992). A 2XCO2 Climate Change Scenario Over Europe Generated Using a 2 . Climate Change Scenario. *Journal of Geophysical Research*, 97(9):11–28.

Giorgi, F. and Mearns, L. O. (1991). Approaches Regional To the Simulation Change : of Climate a Review. *Reviews of Geophysics*, 29(2):191–216.

Glahn, H. R. and Lowry, D. A. (1972). The Use of Model Output Statistics (MOS) in Objective Weather Forecasting.

Glikson, A. Y. (2014). *Evolution of the Atmosphere, Fire and the Anthropocene Climate Event Horizon*. SpringerBriefs in Earth Sciences.

Goly, A., Teegavarapu, R. S. V., and Mondal, A. (2014). Development and evaluation of statistical downscaling models for monthly precipitation. *Earth Interactions*, 18(18):1–28.

Goyal, M. K. and Ojha, C. S. P. (2010). Application of PLS-Regression as Downscaling Tool for Pichola Lake Basin in India. *International Journal of Geosciences*, 1(2):54.

Goyal, S., Kumar, A., Mohapatra, M., Rathore, L. S., Dube, S. K., Saxena, R., and Giri, R. K. (2017). Satellite-based technique for nowcasting of thunderstorms over Indian region. *Journal of Earth System Science*, 126(6):1–13.

Grimes, D., Coppola, E., Verdecchia, M., and Visconti, G. (2003). A Neural Network Approach to Real-Time Rainfall Estimation for Africa Using Satellite Data. *Journal of hydrometeorology*, 4:1119–1133.

Guo, J., Chen, H., Xu, C. Y., Guo, S., and Guo, J. (2012). Prediction of variability of precipitation in the Yangtze River Basin under the climate change conditions based on automated statistical downscaling. *Stochastic Environmental Research and Risk Assessment*, 26(2):157–176.

Gutowski, J. W., Giorgi, F., Timbal, B., Frigon, A., Jacob, D., Kang, H. S., Raghavan, K., Lee, B., Lennard, C., Nikulin, G., O'Rourke, E., Rixen, M., Solman, S., Stephenson, T., and Tangang, F. (2016). WCRP COordinated Regional Downscaling EXperiment (CORDEX): A diagnostic MIP for CMIP6. *Geoscientific Model Development*, 9(11):4087–4095.

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Catherine, A., Bellucci, A., Bao, Q., Chang, P., Corti, S., Fučkar, N. S., Guemas, V., Von Hardenberg, J., Hazeleger, W., Kodama, C., Koenigk, T., Leung, L. R., Lu, J., Luo, J. J., Mao, J., Mizielinski, M. S., Mizuta, R., Nobre, P., Satoh, M., Scoccimarro, E., Semmler, T., Small, J., and Von Storch, J. S. (2016). High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6. *Geoscientific Model Development*, 9(11):4185–4208.

Hagemann, S., Chen, C., Haerter, J. O., Heinke, J., Gerten, D., and Piani, C. (2011). Impact of a Statistical Bias Correction on the Projected Hydrological Changes Obtained from Three GCMs and Two Hydrology Models. *Journal of Hydrometeorology*, 12(4):556–578.

Hall, A. (2005). Calling the shots: The Old English remedy Gif Hors Ofscoten Sie and Anglo-Saxon 'Elf-shot'. *Neuphilologische Mitteilungen*, 106(2):195–209.

Hammami, D., Lee, T. S., Ouarda, T. B., and Le, J. (2012). Predictor selection for downscaling GCM data with LASSO. *Journal of Geophysical Research Atmospheres*, 117:1–11.

Hanel, M., Kožín, R., Heřmanovský, M., and Roub, R. (2017). An R package for assessment of statistical downscaling methods for hydrological climate change impact studies. *Environmental Modelling and Software*, 95:22–28.

Hanssen-Bauer, I., Achberger, C., Benestad, R. E., Chen, D., and Førland, E. J. (2005). Statistical downscaling of climate scenarios over Scandinavia. *Climate Research*, 29(3):255–268.

Hargreaves, J. C. (2010). Skill and uncertainty in climate models. *Wiley Interdisciplinary Reviews: Climate Change*, 1(4):556–564.

Hargreaves, J. C. and Annan, J. D. (2014). Can we trust climate models? *Wiley Interdisciplinary Reviews: Climate Change*, 5(4):435–440.

Hayhoe, K., Wake, C., Anderson, B., Liang, X.-Z., Maurer, E., Zhu, J., Bradbury, J., DeGaetano, A., Stoner, A. M., and Wuebbles, D. (2008). Regional climate change projections for the Northeast USA. *Mitigation and Adaptation Strategies for Global Change*, 13(5-6):425–436.

Hegerl, G. and Zwiers, F. (2011). Use of models in detection and attribution of climate change. *Wiley Interdisciplinary Reviews: Climate Change*, 2(4):570–591.

Hessami, M., Gachon, P., Ouarda, T. B., and St-Hilaire, A. (2008). Automated regression-based statistical downscaling tool. *Environmental Modelling and Software*, 23(6):813–834.

Hewitson, B. C. and Crane, R. G. (1996). Climate downscaling: Techniques and application. *Climate Research*, 7(2):85–95.

Hewitson, B. C. and Crane, R. G. (2006). Consensus between GCM climate change projections with empirical downscaling: Precipitation downscaling over South Africa. *International Journal of Climatology*, 26(10):1315–1337.

Hidalgo, J., Pigeon, G., and Masson, V. (2008). Urban-breeze circulation during the CAPITOUL experiment: Observational data analysis approach. *Meteorology and Atmospheric Physics*, 102(3-4):223–241.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25:1965–1978.

Hillel, D. and Rosenzweig, C. (2011). *Handbook of climate change and agroecosystems : Impacts, adaptation and mitigation*, volume 1.

Huang, Y., Fipps, G., Hotel, R., and Vegas, L. (2003). Modeling Flows in Irrigation Distribution Networks - Model Description and Prototype. *Agricultural Engineering*, 0300(03):1–115.

Huth, R., Kliegrov\´a, S., and Metelka, L. (2008). Non-linearity in statistical downscaling: does it bring an improvement for daily temperature in Europe? *International Journal of Climatology*, 28(December 2007):465–477.

Iizumi, T., Nishimori, M., Dairaku, K., Adachi, S. A., and Yokozawa, M. (2011). Evaluation and intercomparison of downscaled daily precipitation indices over Japan in present-day climate: Strengths and weaknesses of dynamical and bias correction-type statistical downscaling methods. *Journal of Geophysical Research Atmospheres*, 116(1):1–20.

IPCC, I. P. o. C. C. (2014). Overview of MIPs that have applied for CMIP6 Endorsement Overview how MIPs contribute to the three CMIP6 science questions. (December).

Irvine, P. J., Kravitz, B., Lawrence, M. G., and Muri, H. (2016). An overview of the Earth system science of solar geoengineering. *Wiley Interdisciplinary Reviews: Climate Change*, 7(6):815–833.

Jain, S. K. and Kumar, V. (2012). Trend analysis of rainfall and temperature data for India. *Current Science*, 102(1):37–49.

Jaiswal, R. K., Tiwari, H. L., Lohani, A. K., and Yadava, R. N. (2018). Statistical Downscaling of Minimum Temperature of Raipur (C.G.) India. In *Climate Change Impacts Select Proceedings of ICWEES-2016*, pages 35–45.

Jana, C., Alam, N. M., Mandal, D., Shamim, M., and Kaushal, R. (2017). Spatio-temporal rainfall trends in the twentieth century for Bundelkhand region, India. *Journal of Water and Climate Change*, 8(3):441–455.

Janicot, S. and Kamga, F. (2015). Key research and prediction challenges for West and Central Africa. *Report CLIVAR Exchanges (no. 66)*, 19(1):10–15.

Jayadeva, Khemchandani, R., and Chandra, S. (2017). *Twin Support Vector Machines*, volume 659. Springer Nature.

Jena, P., Azad, S., and Rajeevan, M. (2015). Statistical Selection of the Optimum Models in the CMIP5 Dataset for Climate Change Projections of Indian Monsoon Rainfall. *Climate*, 3(4):858–875.

Jeon, S., Paciorek, C. J., and Wehner, M. F. (2015). Quantile-based bias correction and uncertainty quantification of extreme event attribution statements. *Weather and Climate Extremes*, 12:24–32.

Jha, S. K., Mariethoz, G., Evans, J. P., and McCabe, M. F. (2013). Demonstration of a geostatistical approach to physically consistent downscaling of climate modeling simulations. *Water Resources Research*, 49(1):245–259.

Jia, J., Punys, P., and Ma, J. (2012). Hydropower. In Handbook of Climate Change Mitigation.

Jia, P., Zhuang, D., and Wang, Y. (2017). Impacts of temperature and precipitation on the spatiotemporal distribution of water resources in Chinese mega cities: the case of Beijing. *Journal of Water and Climate Change*, page jwc2017038.

Jones, P. G., Thornton, P. K., and Heinke, J. (2009). Generating characteristic daily weather data using downscaled climate model data from the IPCC ' s Fourth Assessment. *Report*, (August):24.

Jones, R., Noguer, M., Hassell, D., Hudson, D., Wilson, S., Jenkins, G., Mitchell, J., Taylor, R., Office, M., and Centre, H. (2004). Generating high resolution climate change scenarios using Precis. *Office*, page 40.

José, R. S. and Pérez, J. L. (2014). Climate Downscaling Processes using WRF / chem and CALMET modelling systems : Madrid case study. In *International Conference on Utility and Cloud Computing*, pages 830–835.

Karl, T. R., Wang, W.-C., Schlesinger, M. E., Knight, R. W., and Portman, D. (1990). A Method of Relating General Circulation Model Simulated Climate to the Observed Local Climate. Part I: Seasonal Statistics. *Journal of Climate*, 3(10):1053–1079.

Kaufmann, R. K., Zhou, L., Tucker, C. J., Slayback, D., Shabanov, N. V., and Myneni, R. B. (2002). IPCC AR5. *Report*, 107(D11):1535.

Keenlyside, N. S. and Ba, J. (2010). Prospects for decadal climate prediction. *Wiley Interdisciplinary Reviews: Climate Change*, 1(5):627–635.

Keller, D. E., Fischer, A. M., Liniger, M. A., Appenzeller, C., and Knutti, R. (2017). Testing a weather generator for downscaling climate change projections over Switzerland. *International Journal of Climatology*, 37(2):928–942.

Khalili, M. and Nguyen, V. T. V. (2018). Efficient Statistical Approach to Multisite Downscaling of Extreme Temperature Series Using Singular-Value Decomposition Technique. *Journal of Hydrologic Engineering*, 23(6):10.

Khalili, M., Van Nguyen, V. T., and Gachon, P. (2013). A statistical approach to multi-site multivariate downscaling of daily extreme temperature series. *International Journal of Climatology*, 33(1):15–32.

Khan, M. S., Coulibaly, P., and Dibike, Y. (2006a). Uncertainty analysis of statistical downscaling methods. *Journal of Hydrology*, 319(1-4):357–382.

Khan, M. S., Coulibaly, P., and Dibike, Y. (2006b). Uncertainty analysis of statistical downscaling methods using Canadian Global Climate Model predictors. *Hydrological Processes*, 20(14):3085–3104.

Kim, J.-S., Son, C.-Y., Moon, Y.-I., and Lee, J.-H. (2017). Seasonal rainfall variability in Korea within the context of different evolution patterns of the central Pacific El Niño. *Journal of Water and Climate Change*, 8(3):412–422.

Kim, J.-W., Chang, J.-T., Baker, N. L., Wilks, D. S., and Gates, W. L. (1984). The Statistical Problem of Climate Inversion: Determination of the Relationship between Local and Large-Scale Climate. *Monthly Weather Review*, 112(10):2069–2077.

Klein, W. and Hammons, G. (1975). Maximum/minimum temperature forecasts based on model output statistics. *Monthly Weather Review*, 103(9):796–806.

Klein, W. H., Lewis, B. M., Crockett, C. W., and Enger, I. (1960). Application of Numerical Prognostic Heights to Surface Temperature Forecasts. *Tellus*, 12(4):378–392.

Klein, W. H., Lewis, B. M., and Enger, I. (1959). Objective prediction of five-day mean temperatures during winter. *Journal of Meteorology*, 16:672–682.

Klein, W. H., Lewis, F., and Casely, G. P. (1967). Automated Nationwide Forecasts of Maximum and Minimum Temperature. *Journal of Applied Meteorology*, 6(4):216–228.

Kossin, J. P., Emanuel, K. A., and Camargo, S. J. (2016). Past and projected changes in western north pacific tropical cyclone exposure. *Journal of Climate*, 29(16):5725–5739.

Krishnamurthy, V. (2011). Extreme Events and Trends in the Indian Summer Monsoon, COLA Technical Report 314. *COLA Technical Report 314 July*, (July):36.

Kumar, V. and Jain, S. K. (2011). Trends in rainfall amount and number of rainy days in river basins of India (1951–2004). *Hydrology Research*, 42(4):290–306.

Kumar, V., Jain, S. K., and Singh, Y. (2010). Analysis of long-term rainfall trends in India. *Hydrological Sciences Journal*, 55(4):484–496.

Lafon, T., Dadson, S., Buys, G., and Prudhomme, C. (2013). Bias correction of daily precipitation simulated by a regional climate model: A comparison of methods. *International Journal of Climatology*, 33(6):1367–1381.

Lakshmanan, V., Gilleland, E., McGovern, A., and Tingley, M. (2015). *Machine Learning and Data Mining Approaches to Climate Science*. Springer International Publishing, 1 edition.

Lawrence, D. M., Hurtt, G. C., Arneth, A., Brovkin, V., Calvin, K. V., Jones, A. D., Jones, C. D., Lawrence, P. J., Noblet-Ducoudré, N. D., Pongratz, J., Seneviratne, S. I., and Shevliakova, E. (2016). The Land Use Model Intercomparison Project (LUMIP) contribution to CMIP6: Rationale and experimental design. *Geoscientific Model Development*, 9(9):2973–2998.

Lee, J. W., Hong, S. Y., Chang, E. C., Suh, M. S., and Kang, H. S. (2014). Assessment of future climate change over East Asia due to the RCP scenarios downscaled by GRIMs-RMP. *Climate Dynamics*, 42(3-4):733–747.

Lee, T. and Jeong, C. (2014). Nonparametric statistical temporal downscaling of daily precipitation to hourly precipitation and implications for climate change scenarios. *Journal of Hydrology*, 510:182–196.

Lenderink, G., Buishand, A., and Van Deursen, W. (2007). Estimates of future discharges of the river Rhine using two scenario methodologies: Direct versus delta approach. *Hydrology and Earth System Sciences*, 11(3):1145–1159.

Lewandowski, C. M. (2015). *Out of Thin Air*. Joseph Henry Press.

Lin, G.-F., Chang, M.-J., and Wu, J.-T. (2017). A Hybrid Statistical Downscaling Method Based on the Classification of Rainfall Patterns. *Water Resources Management*, 31(1):377–401.

Liss, P. S. and Johnson, M. T. (2014). *Ocean-Atmosphere Interactions of Gases and Particles*. Springer Earth System Sciences.

Liu, W., Fu, G., Liu, C., and Charles, S. P. (2013). A comparison of three multi-site statistical downscaling models for daily rainfall in the North China Plain. *Theoretical and Applied Climatology*, 111(3-4):585–600.

Lloyd, E. A. and Winsberg, E. (2018). *Climate Modelling: Philosophical and Conceptual Issues*. Springer Nature.

Lobell, D. B. and Field, C. B. (2007). Global scale climate-crop yield relationships and the impacts of recent warming. *Environmental Research Letters*, 2(1).

Lorenzoni, I., Nicholson-Cole, S., and Whitmarsh, L. (2007). Barriers perceived to engaging with climate change among the UK public and their policy implications. *Global Environmental Change*, 17(3-4):445–459.

Luksch, U. and von Storch, H. (1992). Modeling the Low-Frequency Sea-Surface Temperature Variability in the North Pacific. *Journal of climate*, 5:893–906.

Lyu, S. and Farid, H. (2002). Detecting Hidden Messages Using Higher-Order Statistics and Support Vector Machines. In *5th International Workshop on Information Hiding, The Netherlands,2002*, pages 340–354.

Madsen, H., Lawrence, D., Lang, M., Martinkova, M., and Kjeldsen, T. R. (2014). Review of trend analysis and climate change projections of extreme precipitation and floods in Europe. *Journal of Hydrology*, 519:3634–3650.

Mahmood, R. and Babel, M. S. (2013). Evaluation of SDSM developed by annual and monthly sub-models for downscaling temperature and precipitation in the Jhelum basin, Pakistan and India. *Theoretical and Applied Climatology*, 113(1-2):27–44.

Maity, R., Aggarwal, A., and Chanda, K. (2016). Do CMIP5 models hint at a warmer and wetter India in the 21st century? *Journal of Water and Climate Change*, 7(2):280–295.

Mangasarian, O. L. (2001). Data mining via support vector machines. In *In System Modeling and Optimization*, volume 130, pages 91–112.

Manly, B. F. J. and Alberto, J. A. N. (2017). *Multivariate Statistical Methods A Primer*. Taylor & Francis, New York, fourth edi edition.

Mann, M. E. and Emanuel, K. A. (2006). Atlantic hurricane trends linked to climate change. *Eos, Transactions American Geophysical Union*, 87(24):233–244.

Mann-kendall, D. T. (1987). Mann-Kendall Test For Monotonic Trend. pages 1–4.

Maraun, D. (2013). Bias correction, Quantile mapping, and Downscaling: Revisiting the Inflation Issue. *Journal of Climate*, 27(4):2137–2143.

Maraun, D., Shepherd, T. G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J. M., Hagemann, S., Richter, I., Soares, P. M., Hall, A., and Mearns, L. O. (2017). Towards process-informed bias correction of climate change simulations. In *Nature Climate Change*, volume 7, pages 764–773.

Maraun, D., Wetterhall, F., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust, H. W., Sauter, T., Themeßl, M., Venema, V. K. C., Chun, K. P., Goodess, C. M., Jones, R. G., Onof, C., Vrac, M., and Thiele-Eich, I. (2010). Precipitation downscaling under climate change: Recent developements to bridge the gap between dynamical models and the end user. *Reviews of Geophysics*, 48(2009RG000314):1–38.

Maraun, D. and Widmann, M. (2018). *Statistical Downscaling and Bias Correction for Climate Research*. Cambridge University Press, New York,USA; Port Melbourne,Australia; New Delhi, India; Singapore, first edition.

Marshak, A. and Davis, A. (2005). *3D Radiative Transfer in Cloudy Atmospheres*. Springer-Verlag.

Maurer, E. P., Hidalgo, H. G., Das, T., Dettinger, M. D., and Cayan, D. R. (2010). The utility of daily large-scale climate data in the assessment of climate change impacts on daily streamflow in California. *Hydrology and Earth System Sciences*, 14(6):1125–1138.

McGarigal, K., Compton, B., Plunkett, E., Deluca, B., and Grand, J. (2017). Designing Sustainable Landscapes : Climate data A project of the University of Massachusetts Landscape Ecology Lab. Technical Report March.

Mcginnis, S., Nychka, D., and Mearns, L. O. (2014). A new distribution mapping technique for climate model bias correction. *4th International Workshop on Climate Informatics*.

McGuffie, K. and Henderson-Sellers, A. (2005). *A Climate Modelling Primer*.

Mearns, L. O., Bukovsky, M. S., Pryor, S. C., and Magaña, V. (2018). Downscaling of climate information. In *Climate Modelling: Philosophical and Conceptual Issues*, pages 199–269. Springer International Publishing Switzerland.

Meenu, R., Rehana, S., and Mujumdar, P. P. (2013). Assessment of hydrologic impacts of climate change in Tunga–Bhadra river basin, India with HEC-HMS and SDSM. *Hydrological Processes*, 27(11):1572–1589.

Mehrotra, R. and Sharma, A. (2010). Development and application of a multisite rainfall stochastic downscaling framework for climate change impact assessment. *Water Resources Research*, 46(7):1–17.

Meinshausen, M., Vogel, E., Nauels, A., Lorbacher, K., Meinshausen, N., Etheridge, D. M., Fraser, P. J., Montzka, S. A., Rayner, P. J., Trudinger, C. M., Krummel, P. B., Beyerle, U., Canadell, J. G., Daniel, J. S., Enting, I. G., Law, R. M., Lunder, C. R., O'Doherty, S., Prinn, R. G., Reimann, S., Rubino, M., Velders, G. J., Vollmer, M. K., Wang, R. H., and Weiss, R. (2017). Historical greenhouse gas concentrations for climate modelling (CMIP6). *Geoscientific Model Development*, 10(5):2057–2116.

Melnikova, I. N. and Vasilyev, A. V. (2005). *Short-Wave Solar Radiation in the Earth's Atmosphere*. Springer.

Mendes, D., Marengo, J. A., Rodrigues, S., and Oliveira, M. (2014). Downscaling Statistical Model Techniques for Climate Change Analysis Applied to the Amazon Region. *Advances in Artificial Neural Systems*, 2014(595462):1–10.

MICHEL, K. (1974). *Impact of Aerospace Technology on Studies of the Earth's Atmosphere*.

Mitra, A., Banerjee, K., and Sett, S. (2012). Spatial variation in organic carbon density of mangrove soil in Indian Sundarbans. *National Academy Science Letters*, 35(3):147–154.

Mizukami, N., Clark, M. P., Gutmann, E. D., Mendoza, P. A., Newman, A. J., Nijssen, B., Livneh, B., Hay, L. E., Arnold, J. R., and Brekke, L. D. (2016). Implications of the Methodological Choices for Hydrologic Portrayals of Climate Change over the Contiguous United States: Statistically Downscaled Forcing Data and Hydrologic Models. *Journal of Hydrometeorology*, 17(1):73–98.

Model, C., Project, I., Climate, C., Cycle, C., Intercomparison, M., Feedback, C., Intercomparison, M., Model, A., Project, I., Climate, D., Project, P., Model, G., Project, I., Monsoons, G., Intercomparison, M., Resolution, H., Intercomparison, M., Surface, L., Moisture, S., Model, O., Project, I., Driver, P., Model, R., Project, I., Modelling, P., Project, I., Forcing, R., Intercomparison, M., Model, S., Project, I., Model, S., Project, I., Forcings, V., Intercomparison, M., Regional, C., and Downscaling, C. (2014). Application for CMIP6-Endorsed MIPs. (April).

Mondal, A. and Mujumdar, P. P. (2015). On the detection of human influence in extreme precipitation over India. *Journal of Hydrology*, 529(September):1161–1172.

Mpelasoka, F. S. and Chiew, F. H. S. (2009). Influence of Rainfall Scenario Construction Methods on Runoff Projections. *Journal of Hydrometeorology*, 10(5):1168–1183.

Murphy, J. (2000). Predictions of climate change over Europe using statistical and dynamical downscaling techniques. *International Journal of Climatology*, 20(5):489–501.

Murphy, K. M., Shleifer, A., and Vishny, R. W. (1989). Industrialization and the Big Push Authors. *Journal of Political Economy*, 97(5):1003–1026.

Murty, M. and Raghava, R. (2016). *Support Vector Machines and Perceptrons*.

Name, L. (2016). CMIP6 Data Citation GUI User Guide Table of Contents. 2.

National Institute of Hydrology (2015). Snow and Glacier Contribution in the Satluj River at Bhakra Dam. *Report*, 667:1–19.

National Research Council (2007). *Understanding Multiple Environmental Stresses*.

Notz, D. (2012). Challenges in simulating sea ice in Earth System Models. *Wiley Interdisciplinary Reviews: Climate Change*, 3(6):509–526.

O'Connor, R. E., Bord, R. J., and Fisher, A. (1999). Risk perceptions, general environmental beliefs, and willingness to address climate change. *Risk Analysis*, 19(3):461–471.

Olsen, R. and Gilroy, K. (2012). Risk Informed Decision Making in a Changing Climate.

Onyutha, C., Tabari, H., Rutkowska, A., Nyeko-Ogiramoi, P., and Willems, P. (2016). Comparison of different statistical downscaling methods for climate change rainfall projections over the Lake Victoria basin considering CMIP3 and CMIP5. *Journal of Hydro-Environment Research*, 12:31–45.

Open Geospatial Consortium (2015). Geographic information - Well known text representation of coordinate reference systems. Technical report.

Orlowsky, B., Bothe, O., Fraedrich, K., Gerstengarbe, F. W., and Zhu, X. (2010). Future climates from bias-bootstrapped weather analogs: An application to the Yangtze River basin. *Journal of Climate*, 23(13):3509–3524.

Pagé, C., Terray, L., and Boé, J. (2009). dsclim: A software package to downscale climate scenarios at regional scale using a weather-typing based statistical methodology. *Methodology*, page 58.

Pahlavan, H. A., Zahraie, B., Nasseri, M., and Mahdipour Varnousfaderani, A. (2017). Improvement of multiple linear regression method for statistical downscaling of monthly precipitation. *International Journal of Environmental Science and Technology*, page 16.

Pal, I. and Al-Tabbaa, A. (2009). Trends in seasonal precipitation extremes - An indicator of 'climate change' in Kerala, India. *Journal of Hydrology*, 367(1-2):62–69.

Panel, C. and Eyring, V. (2015). Application for CMIP6-Endorsed MIPs. (November 2014).

Pare, T. P. (2017). Biosciences and Plant Biology. *Int. J. Curr. Res. Biosci. Plant Biol*, 4(6):42–50.

Parrella, F. and Parrella, F. (2007). Online Support Vector Regression. (June).

Parvaze, S., Parvaze, S., Haroon, S., Khurshid, N., Khan, J. N., and Ahmad, L. (2016). Projected Change in Climate Under A2 Scenario in Dal Lake Catchment Area of Srinagar City in Jammu and Kashmir. *Current World Environment*, 11(2):429–438.

Patz, J. A., Campbell-Lendrum, D., Holloway, T., and Foley, J. A. (2005). Impact of regional climate change on human health. *Nature*, 438:310–317.

Paul, A., Bhowmik, R., Chowdary, V. M., Dutta, D., Sreedhar, U., and Sankar, H. R. (2017). Trend analysis of time series rainfall data using robust statistics. *Journal of Water and Climate Change*, page 11.

Piani, C., Haerter, J. O., and Coppola, E. (2010). Statistical bias correction for daily precipitation in regional climate models over Europe. *Theoretical and Applied Climatology*, 99(1-2):187–192.

Poggio, L. and Gimona, A. (2015). Downscaling and correction of regional climate models outputs with a hybrid geostatistical approach. *Spatial Statistics*, 14:4–21.

Pont, F. J. (2014). *Alien Skies : Planetary Atmospheres from Earth to Exoplanets*. Springer New York.

Puma, M. J. (2012). A Holistic Approach to Guide Development of Future Climate Scenarios for Water Resource Applications. *Journal of Contemporary Water Research & Education*, 147(3):41–48.

Quintana Seguí, P., Ribes, A., Martin, E., Habets, F., and Boé, J. (2010). Comparison of three downscaling methods in simulating the impact of climate change on the hydrology of Mediterranean basins. *Journal of Hydrology*, 383(1-2):111–124.

Racsko, P., Szeidl, L., and Semenov, M. (1991). A serial approach to local stochastic weather models. *Ecological Modelling*, 57(1-2):27–41.

Rafferty, J. P. (2011). *Dynamic Earth: Storms, Violent Winds And Earth's Atmosphere.* Encyclopædia Britannica.

Raje, D. and Mujumdar, P. P. (2011). A comparison of three methods for downscaling daily precipitation in the Punjab region. *Hydrological Processes*, 25(23):3575–3589.

Raju, K. S. and Kumar, D. N. (2014). Ranking of global climate models for India using multicriterion analysis. *Climate Research*, 60(2):103–117.

Rao, V. B., Franchito, S. H., Gerólamo, R. O. P., Giarolla, E., Ramakrishna, S. S. V. S., Rao, B. R. S., and Naidu, C. V. (2016). Himalayan Warming and Climate Change in India. *American Journal of Climate Change*, 05(04):558–574.

Refsgaard, J. C., Madsen, H., Andréassian, V., Arnbjerg-Nielsen, K., Davidson, T. A., Drews, M., Hamilton, D. P., Jeppesen, E., Kjellström, E., Olesen, J. E., Sonnenborg, T. O., Trolle, D., Willems, P., and Christensen, J. H. (2014). A framework for testing the ability of models to project climate change and its impacts. *Climatic Change*, 122(1-2):271–282.

Richardson, C. W. C. (1981). Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research*, 17(1):182.

Roy, S. S. and Balling, R. C. (2004). Trends in extreme daily precipitation indices in India. *International Journal of Climatology*, 24(4):457–466.

Rummukainen, M. (1997). Methods for statistical downscaling of GCM simulations. *SMHI Rapporter. Meteorologi och Klimatologi*, (80).

Rummukainen, M. (2010). State-of-the-art with regional climate models. *Wiley Interdisciplinary Reviews-Climate Change*, 1(1):82–96.

Rummukainen, M. (2012). Changes in climate and weather extremes in the 21st century. *Wiley Interdisciplinary Reviews: Climate Change*, 3(2):115–129.

Rupa Kumar, K., Sahai, A. K., Krishna Kumar, K., Patwardhan, S. K., Mishra, P. K., Revadekar, J. V., and Kamala, K. (2006). High-resolution climate change scenarios for India for the 21stcentury. *Current Science*, 90(3):334–345.

Ryu, J. H., Palmer, R. N., Wiley, M. W., and Jeong, S. (2009). Mid-range streamflow forecasts based on climate modeling - Statistical correction and evaluation. *Journal of the American Water Resources Association*, 45(2):355–368.

Sa'adi, Z., Shahid, S., Chung, E. S., and bin Ismail, T. (2017). Projection of spatial and temporal changes of rainfall in Sarawak of Borneo Island using statistical downscaling of CMIP5 models. *Atmospheric Research*, 197(August):446–460.

Sachindra, D. A., Huang, F., Barton, A., and Perera, B. J. C. (2014a). Statistical downscaling of general circulation model outputs to precipitation-part 1: Calibration and validation. *International Journal of Climatology*, 34(11):3264–3281.

Sachindra, D. A., Huang, F., Barton, A., and Perera, B. J. C. (2014b). Statistical downscaling of general circulation model outputs to precipitation-part 2: Bias-correction and future projections. *International Journal of Climatology*, 34(11):3282–3303.

Saha, K. (2008). *The Earth's Atmosphere : Its Physics and Dynamics*. Springer-Verlag.

Salathe, E. P., Mote, P. W., and Wiley, M. W. (2007). Review of scenario selection and downscaling methods for the assessment of climate change impacts on hydrology in the United States pacific northwest.

Salvi, K., Kannan, S., and Ghosh, S. (2011). Statistical downscaling and bias-correction for projections of Indian rainfall and temperature in climate change studies. *4th International Conference on . . .*, 19(6):7–11.

Santos, A. P. P., Pinto Junior, O., Santos, S. R. Q., Lima, F. J. L., Souza, E. B., Morais, A. A. R., Avila, E. E., and Pedernera, A. (2017). Climatic Projections of Lightning in Southeastern Brazil Using CMIP5 Models in RCP ' s Scenarios 4 . 5 and 8 . 5. *American Journal of Climate Change*, 6:539–553.

Sarhadi, A., Burn, D. H., Johnson, F., Mehrotra, R., and Sharma, A. (2016a). Water resources climate change projections using supervised nonlinear and multivariate soft computing techniques. *JOURNAL OF HYDROLOGY*, 536:119–132.

Sarhadi, A., Burn, D. H., Yang, G., and Ghodsi, A. (2016b). Advances in projection of climate change impacts using supervised nonlinear dimensionality reduction techniques. *Climate Dynamics*, 48(3-4):1329–1351.

Sarr, A. (2017). Multi-Scale Characteristics of Precipitation and Temperature over West Africa Using SMHI-RCA Driven by GCMs under RCP8.5. *American Journal of Climate Change*, 06(03):455–486.

Scheuerer, M. (2014). Generating Calibrated Ensembles of Physically Realistic , High-Resolution Precipitation Forecast Fields Based on GEFS Model Output. Technical report, University of Colorado, Cooperative Institute for Research in Environmental Sciences, and NOAA.

Scheuerer, M. (2018). Appendix a : Detailed description of the quantile mapping procedure.

Schneider, B. (2012). Climate model simulation visualization from a visual studies perspective. *Wiley Interdisciplinary Reviews: Climate Change*, 3(2):185–193.

Semenov, M. A. and Brooks, R. J. (1999). Spatial interpolation of the LARS-WG stochastic weather generator in Great Britain. *Climate Research*, 11(2):137–148.

Sendhil, R., Ramasundaram, P., Pal, M. R., Thimmappa, K., and Sharma, I. (2016). Tracking the Yield Sensitivity of Rice–Wheat System to Weather Anomalies. *National Academy Science Letters*, 39(6):401–405.

Serur, A. B. and Sarma, A. K. (2017). Statistical Downscaling of Daily Temperatures and Precipitation Data from Coupled Model Inter-comparison Project 5 (CMIP5)-RCPs Experiment: In Weyib River Basin, Southeastern Ethiopia. In *Climate Change Impacts, Water Science*, pages 47–70. Springer Nature.

Shah, S. K., Bhattacharyya, A., and Chaudhary, V. (2007). Reconstruction of June-September precipitation based on tree-ring data of teak (Tectona grandis L.) from Hoshangabad, Madhya Pradesh, India. *Dendrochronologia*, 25(1):57–64.

Sharma, A. K., Singh, S. K., Kulkarni, A. V., and Ajai (2013). Glacier inventory in Indus, Ganga and Brahmaputra basins of the Himalaya. *National Academy Science Letters*, 36(5):497–505.

Sharma, D., Gupta, A. D., and Babel, M. S. (2007). Spatial disaggregation of bias-corrected GCM precipitation for improved hydrologic simulation: Ping River Basin, Thailand. *Hydrology and Earth System Sciences*, 11(4):1373–1390.

Sharma, S., Couturier, S., and Côté, S. D. (2009). Impacts of climate change on the seasonal distribution of migratory caribou. *Global Change Biology*, 15(10):2549–2562.

Shaw, G. H. (2016). *Earth's Early Atmosphere and Oceans, and The Origin of Life*, volume 53.

Shrestha, A. B., Wake, C. P., Mayewski, P. A., and Dibb, J. E. (2000). Precipitation fluctuations in the Himalaya and its vicinity: An analysis based on temperature records from Nepal. *International Journal of Climate*, 20(9):317–327.

Shrestha, S., Babel, M. S., and Pandey, V. P. (2014). *Climate Change and Water Resources*. CRC Press Taylor & Francis Group.

Siberia, P. E., Onishchuk, N., Khodzher, T., and Netsvetaeva, O. (2017). Chemical Composition of Atmospheric Precipitation at Monitoring Sites in. *American Journal of Climate Change*, 6:374–384.

Sigdel, M. and Ma, Y. (2016). Evaluation of future precipitation scenario using statistical downscaling model over humid, subhumid, and arid region of Nepal—a case study. *Theoretical and Applied Climatology*, 123(3-4):453–460.

Singh, D., Gupta, R. D., and Jain, S. K. (2013). Study of Long-Term Trend in River Discharge of Sutlej River (N-W Himalayan Region). *Environment*, page 11.

Singh, D., Gupta, R. D., and Jain, S. K. (2015a). Assessment of impact of climate change on water resources in a hilly river basin. *Arabian Journal of Geosciences*, 8(12):10625–10646.

Singh, D., Gupta, R. D., and Jain, S. K. (2015b). Statistical analysis of long term spatial and temporal trends of temperature parameters over Sutlej river basin, India. *Journal of Earth System Science*, 124(1):17–35.

Singh, D., Jain, S., and Gupta, R. D. (2015c). Global Climate Model : A Comprehensive Tool In Climate Change Impact Studies. *Environment*, 02(08):15.

Singh, D., Jain, S. K., and Gupta, R. D. (2012). Evaluating Impacts of River Flooding on Human Society with Special Reference to Indian Monsoon. In *3rd International Symposium on Response of Asian Rivers to Climate Change : Past, Present and Future Scenario*, page 7, Hyderabad, India.

Singh, D., Jain, S. K., and Gupta, R. D. (2015d). Statistical downscaling and projection of future temperature and precipitation change in middle catchment of Sutlej River Basin, India. *Journal of Earth System Science*, 124(4):843–860.

Singh, D., Jain, S. K., and Gupta, R. D. (2015e). Trend d in Obs served and a Pro ojected Maxim mum and d Minim mum Temp perature e over N-W Him malaya n Basin n. *Journal of Mountain Sci ence*, 12(2):417–433.

Singh, D., Jain, S. K., Gupta, R. D., Kumar, S., Rai, S., and Jain, N. (2016a). Analyses of Observed and Anticipated Changes in Extreme Climate Events in the Northwest Himalaya. *Climate*, 4(1):9.

Singh, M. K., Gupta, R. D., Snehmani, Bhardwaj, A., and Ganju, A. (2016b). Effect of sensor modelling methods on computation of 3-D coordinates from Cartosat-1 stereo data. *Geocarto International*, 31(5):506–526.

Singh, M. K., Gupta, R. D., Snehmani, Bhardwaj, A., and Ganju, A. (2016c). Scenario-Based Validation of Moderate Resolution DEMs Freely Available for Complex Himalayan Terrain. *Pure and Applied Geophysics*, 173(2):463–485.

Singh, M. K., Gupta, R. D., Snehmani, Kumar, S., and Ganju, A. (2016d). Assessment of freely available CartoDEM V1 and V1.1R1 with respect to high resolution aerial photogrammetric DEM in high mountains. *Geocarto International*, 31(9):943–955.

Singh, M. K., Snehmani, Gupta, R. D., Bhardwaj, A., Joshi, P. K., and Ganju, A. (2015f). High resolution DEM generation for complex snow covered Indian Himalayan Region using ADS80 aerial push-broom camera: a first time attempt. *Arabian Journal of Geosciences*, 8(3):1403–1414.

Smerdon, J. E. (2012). Climate models as a test bed for climate reconstruction methods: Pseudoproxy experiments. *Wiley Interdisciplinary Reviews: Climate Change*, 3(1):63–77.

Smid, M. and Costa, A. C. (2017). Climate projections and downscaling techniques: a discussion for impact studies in urban systems. *International Journal of Urban Sciences*, 0(0):1–31.

Smith, S., Zhou, Y., Kyle, P., Wang, H., and Yu, H. (2015). A Community Emissions Data System (CEDS): Emissions for CMIP6 and beyond. Technical Report April.

Smitha, P. S., Narasimhan, B., Sudheer, K. P., and Annamalai, H. (2018). An improved bias correction method of daily rainfall data using a sliding window technique for climate change impact assessment. *Journal of Hydrology*, 556:100–118.

Smola, A. J. and Sch, B. (2003). A Tutorial on Support Vector Regression.

Snehmani, Singh, M. K., Gupta, R. D., Bhardwaj, A., and Joshi, P. K. (2015). Remote sensing of mountain snow using active microwave sensors: a review. *Geocarto International*, 30(1):1–27.

Sonali, P. and Nagesh Kumar, D. (2016). Detection and attribution of seasonal temperature changes in India with climate models in the CMIP5 archive. *Journal of Water and Climate Change*, 7(1):83–102.

Srinivasa Raju, K., Sonali, P., and Nagesh Kumar, D. (2017). Ranking of CMIP5-based global climate models for India using compromise programming. *Theoretical and Applied Climatology*, 128(3-4):563–574.

Stennett-Brown, R. K., Jones, J. J., Stephenson, T. S., and Taylor, M. A. (2017). Future Caribbean temperature and rainfall extremes from statistical downscaling. *International Journal of Climatology*, 37(14):4828–4845.

Stephens, E. M., Edwards, T. L., and Demeritt, D. (2012). Communicating probabilistic information from climate model ensembles-lessons from numerical weather prediction. *Wiley Interdisciplinary Reviews: Climate Change*, 3(5):409–426.

Stoean, C. and Stoean, R. (2014). *Support Vector Machines and Evolutionary Algorithms for Classifi cation. Single or Together.*

Stoner, A. M. K., Hayhoe, K., Yang, X., and Wuebbles, D. J. (2013). An asynchronous regional regression model for statistical downscaling of daily climate variables. *International Journal of Climatology*, 33(11):2473–2494.

Stott, P. A., Christidis, N., Otto, F. E., Sun, Y., Vanderlinden, J. P., van Oldenborgh, G. J., Vautard, R., von Storch, H., Walton, P., Yiou, P., and Zwiers, F. W. (2016). Attribution of extreme weather and climate-related events. *Wiley Interdisciplinary Reviews: Climate Change*, 7(1):23–41.

Stouffer, R. J., Eyring, V., Meehl, G. A., Bony, S., Senior, C., Stevens, B., and Taylor, K. E. (2017). CMIP5 scientific gaps and recommendations for CMIP6. *Bulletin of the American Meteorological Society*, 98(1):95–105.

Sun, F., Walton, D. B., and Hall, A. (2015). A hybrid dynamical-statistical downscaling technique. Part II: End-of-Century warming projections predict a new climate state in the Los Angeles region. *Journal of Climate*, 28(12):4618–4636.

Surampalli, R. and Zhang, T. C. (2013). *Climate Change Modeling , Mitigation , and Adaptation*.

Suykens, J. and Leuven, K. U. (2003). A (short) Introduction to Support Vector Machines and Kernelbased Learning. (April).

Tang, J., Niu, X., Wang, S., Gao, H., Wang, X., and Wu, J. (2016a). Statistical downscaling and dynamical downscaling of regional climate in China: Present climate evaluations and future climate projections. *Journal of Geophysical Research*, 121(5):2110–2129.

Tang, J., Niu, X., Wang, S., Gao, H., Wang, X., and Wu, J. (2016b). Statistical downscaling and dynamical downscaling of regional climate in China: Present climate evaluations and future climate projections. *Journal of Geophysical Research : Atmospheres*, 121(5):2110–2129.

Tawde, S. A. (2013). *Investigation of orographically induced rainfall over Western Ghats and its association with other monsoon parameters*. PhD thesis.

Taxak, A. K., Murumkar, A. R., and Arya, D. S. (2014). Long term spatial and temporal rainfall trends and homogeneity analysis in Wainganga basin, Central India. *Weather and Climate Extremes*, 4:50–61.

Terink, W., Hurkmans, R. T. W. L., Torfs, P. J. J. F., and Uijlenhoet, R. (2010). Evaluation of a bias correction method applied to downscaled precipitation and temperature reanalysis data for the Rhine basin. *Hydrology and Earth System Sciences*, 14(4):687–703.

Teutschbein, C. and Seibert, J. (2012). Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *Journal of Hydrology*, 456-457:12–29.

Teutschbein, C., Wetterhall, F., and Seibert, J. (2011). Evaluation of different downscaling techniques for hydrological climate-change impact studies at the catchment scale. *Climate Dynamics*, 37(9-10):2087–2105.

Thakural, L. N., Kumar, S., Jain, S. K., and Ahmad, T. (2018). *The Impact of Climate Change on Rainfall Variability: A Study in Central Himalayas*.

Timmreck, C. (2012). Modeling the climatic effects of large explosive volcanic eruptions. *Wiley Interdisciplinary Reviews: Climate Change*, 3(6):545–564.

Toggweiler, J. and Key, R. (2001). Ocean circulation: Thermohaline circulation. *Encyclopedia of Atmospheric Sciences*, 4(December 2007):1549–1555.

Toggweiler, J. and Key, R. (2009). The annual cycle of heavy precipitation across the United Kingdom: a model based on extreme value statistics. *International Journal of Climatology*, 29(12):1731–1744.

Tripathi, S., Srinivas, V., and Nanjundiah, R. (2006). Downscaling of precipitation for climate change scenarios: a support vector machine approach. In *Journal of Hydrology*, volume 330, pages 621–640.

Tselioudis, G., Douvis, C., and Zerefos, C. (2012). Does dynamical downscaling introduce novel information in climate model simulations of precipitation change over a complex topography region? *International Journal of Climatology*, 32(10):1572–1578.

Twumasi, Y. A., Merem, E. C., Ayala-silva, T., Osei, A., Petja, B. M., and Alexander, K. (2017). Techniques of Remote Sensing and GIS as Tools for Visualizing Impact of Climate Change-Induced Flood in the Southern African Region. *American Journal of Climate Change*, 6:306–327.

UNDP (2011). Preparing Low-Emission Climate-Resilient Development Strategies.

Urich, P., Kouwenhoven, P., Freas, K., and Van Der Tak, L. (2014). New IPCC climate models released: Understanding the planning implications for water resiliency. Technical Report 6.

Valley, T. (2015). Between the Lines : A GIS Based Approach to Reading the Spatial and Topographical Relationships between Cursus Monuments in the Upper. (September).

Vandal, T., Kodra, E., Ganguly, S., Michaelis, A., Nemani, R., and Ganguly, A. R. (2017). DeepSD : Generating High Resolution Climate Change Projections through Single Image Super-Resolution. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, volume 4, pages 1663–1672.

Vogt, G. L. and D., E. (2006). *Earth's Outer Atmosphere : Bordering Space.* Twenty-First Century Books.

von Storch, H. (2010). Climate models and modeling: an editorial essay. *Wiley Interdisciplinary Reviews: Climate Change*, 1(3):305–310.

von Storch, H. and Navarra, A. (1995). *Analysis of Climate Variability: Application of Statistical Techniques*, volume 1.

von Storch, H. and Navarra, A. (1999). *Analysis of Climate Variability.*

Von Storch, H., Zorita, E., and Cubasch, U. (1993). Downscaling of global climate change estimates to regional scales: an application to Iberian rainfall in wintertime. *Journal of Climate*, 6(6):1161–1171.

von Storch, H. and Zwiers, F. W. (1999). *Statistical Analysis in Climate Research*, volume 95.

Von Storch, J. S., Kharin, V. V., Cubasch, U., Hegerl, G. C., Schriever, D., Von Storch, H., and Zorita, E. (1997). A description of a 1260-year control integration with the coupled ECHAM1/LSG general circulation model. *Journal of Climate*, 10(7):1525–1543.

Vu, M. T., Aribarg, T., Supratid, S., Raghavan, S. V., and Liong, S. Y. (2016). Statistical downscaling rainfall using artificial neural network: significantly wetter Bangkok? *Theoretical and Applied Climatology*, 126(3-4):453–467.

Walton, D. B., Hall, A., Berg, N., Schwartz, M., and Sun, F. (2017). Incorporating snow albedo feedback into downscaled temperature and snow cover projections for California's Sierra Nevada. *Journal of Climate*, 30(4):1417–1438.

Walton, D. B., Sun, F., Hall, A., and Capps, S. (2015). A hybrid dynamical-statistical downscaling technique. Part I: Development and validation of the technique. *Journal of Climate*, 28(12):4597–4617.

Wang, J., Nathan, R., Horne, A., Peel, M. C., Wei, Y., and Langford, J. (2017). Evaluating four downscaling methods for assessment of climate change impact on ecological indicators. *Environmental Modelling & Software*, 96:68–82.

Werner, A. T. and Cannon, A. J. (2016). Hydrologic extremes - An intercomparison of multiple gridded statistical downscaling methods. *Hydrology and Earth System Sciences*, 20(4):1483–1508.

Wigley, T. and Barnett, T. (1990). Detection of the greenhouse effect in the observations. In *Climate Change: The IPCC Scientific Assessment*, pages 239–256.

Wigley, T. M. L. and Kelly, P. M. (1990). Holocene climatic change, 14C wiggles and variations in solar irradiance. 330:547–560.

Wigley, T., P., Jones, D., Briffa, K. R., and Smith, G. (1990). Obtaining subgrid scale information from coarse-resolution general circulation model output. *J. Geophys. Res.*, 95(1):1943–1953.

Wikipedia (2015). General Circulation Model.

Wilby, R. L., Charles, S. P., Zorita, E., Timbal, B., Whetton, P., and Mearns, L. O. (2004). Guidelines for Use of Climate Scenarios Developed from Statistical Downscaling Methods. *Analysis*, 27(August):1–27.

Wilby, R. L. and Dawson, C. W. (2013). The statistical downscaling model: Insights from one decade of application. *International Journal of Climatology*, 33(7):1707–1719.

Wilby, R. L., Dawson, C. W., and Barrow, E. M. (2002). Sdsm — a Decision Support Tool for the Assessment of Regional Climate Change Impacts. *Environ. Model. Softw.*, 17(2):145–157.

Wilby, R. L., Tomlinson, O. J., and Dawson, C. W. (2003). Multi-site simulation of precipitation by conditional resampling. *Climate Research*, 23(3):183–194.

Wilby, R. L. and Wigley, T. M. L. (1997). Downscaling general circulation model output: a review of methods and limitations. *Prog. Phys.*, 21:530–548.

Wilby, R. L. and Wigley, T. M. L. (2000). Precipitation predictors for downscaling: Oberserved and General Circulation Model relationships. *Int. J. Climatol.*, 20:641–661.

Wilby, R. L., Wigley, T. M. L., Conway, D., Jones, P. D., Hewitson, B. C., Main, J., and Wilks, D. S. (1998). Statistical downscaling of general circulation model output: A comparison of methods. *Water Res. Research.*, 34(11):2995–3008.

Wilks, D. S. (1992). Adapting stochastic weather generation algorithms for climate change studies. *Climatic Change*, 22(1):67–84.

Willems, P. and Olsson, J. (2009). Climate Change Impact Assessment on Urban Rainfall Extremes and Urban Drainage : Methodologies and. *Change*, 6(1):149–154.

Wood, A. W., Leung, L. R., Sridhar, V., and Lettenmaier, D. P. (2004). Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Climatic Change*, 62(1-3):189–216.

Xu, C.-y. (1999). From GCMs to river flow: a review of downscaling methods and hydrologic modelling approaches. *PPT*, 23(2):229–249.

Yan, Z., Bate, S., Chandler, R. E., Isham, V., and Wheater, H. (2002). An Analysis of Daily Maximum Wind Speed in Northwestern Europe Using Generalized Linear Models. *American Meterological Society*, 15:2073–2088.

Yang, C., Wang, N., and Wang, S. (2017a). A comparison of three predictor selection methods for statistical downscaling. *International Journal of Climatology*, 37(3):1238–1249.

Yang, X. J., Xu, Z. X., Liu, W. F., and Liu, L. (2017b). Spatiotemporal characteristics of extreme precipitation at multiple timescales over Northeast China during 1961–2014. *Journal of Water and Climate Change*, 8(3):535–556.

Yatagai, A., Kamiguchi, K., Arakawa, O., Hamada, A., Yasutomi, N., and Kitoh, A. (2012). Aphrodite constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges. *Bulletin of the American Meteorological Society*, 93(9):1401–1415.

Yehia, A. G., Fahmy, K. M., Mehany, M. A. S., and Mohamed, G. G. (2017). Impact of extreme climate events on water supply sustainability in Egypt: case studies in Alexandria region and Upper Egypt. *Journal of Water and Climate Change*, 8(3):484–494.

Yhang, Y. B., Sohn, S. J., and Jung, I. W. (2017). Application of Dynamical and Statistical Downscaling to East Asian Summer Precipitation for Finely Resolved Datasets. *Advances in Meteorology*, 2017:9.

Yiğit, E. (2015). *Atmospheric and Space Sciences : Neutral Atmospheres*, volume 1. SpringerBriefs in Earth Sciences.

Yu, L., Wang, S., Lai, K. K., and Zhou, L. (2008). *Bio-inspired credit risk analysis computational intelligence with support vector machines*.

Yu, Z., Luo, Y., Yang, K., and Deng, Q. (2017). Analysis on the Climate Change Characteristics of Dianchi Lake Basin under the Background of Global Warming. In *Earth and Environmental Science*, volume 63, page 8.

Zhang, S., Chen, F., He, X., and Liu, B. (2017). A new downscaling approach and its performance with bias correction and spatial disaggregation as contrast. *Journal of Water and Climate Change*, page jwc2017010.

Zhou, T., Turner, A. G., Kinter, J. L., Wang, B., Qian, Y., Chen, X., Wu, B., Liu, B., Zou, L., and He, B. (2016). GMMIP (v1.0) contribution to CMIP6: Global Monsoons Model Inter-comparison Project. *Geoscientific Model Development*, 9(10):3589–3604.

Zorita, E. and von Storch, H. (1999). The Analog Method as a Simple Statistical Downscaling Technique: Comparison with More Complicated Methods. *Journal of Climate*, 12:2474–2489.